

String.Latin+ 1.2

**Zeichen in Unicode für die elektronische Verarbeitung von
Namen und den Datenaustausch in Europa**

Fachgruppe String.Latin

Veröffentlicht 15. Januar 2020

String.Latin+ 1.2: Zeichen in Unicode für die elektronische Verarbeitung von Namen und den Datenaustausch in Europa

PDF generiert am 15.1.2020 7:14 Uhr.

Veröffentlicht 15. Januar 2020

Copyright © 2020 KoSIT —Koordinierungsstelle für IT Standards, Bremen [<http://www.xoev.de>]

Zusammenfassung

Dieses Dokument ist eine kommentierte und ergänzte Fassung der DIN SPEC 91379 vom März 2019, mit der eine in Deutschland relevante Teilmenge des Standards Unicode festgelegt wird. Alle Aussagen und Festlegungen der DIN SPEC 91379 sind unverändert auch in diesem Dokument enthalten. Dies gilt insbesondere für die Festlegungen zu Schriftzeichengruppen und Datentypen sowohl im normativen, als auch im erweiterten Bereich.

Die in diesem Dokument zusätzlich enthaltenen Kommentare und Ergänzungen sollen Nutzern der DIN SPEC 91379, die überwiegend Behörden und Organisationseinheiten des öffentlich-rechtlichen Bereichs sein werden, Hinweise zu deren Anwendung geben. So enthält dieses Dokument einen Abschnitt über die Beschlusslage der öffentlichen Verwaltung, die Behörden unter bestimmten Umständen zur Anwendung der DIN SPEC 91379 verpflichtet (Abschnitt G), und eine Liste häufig gestellter Fragen und Antworten (Abschnitt E). Zudem wird die Bedeutung der Unicode Normalform NFC für die Interoperabilität bei Datenübermittlungen, die auf der DIN SPEC 91379 basieren, in Abschnitt B.7 dargestellt.

Inhaltlich ist die DIN SPEC 91379 ein Nachfolger des von der KoSIT im Auftrag des IT Planungsrats herausgegebenen Standards „Lateinische Zeichen in Unicode“. Dieser ist in der Version 1.1 vom IT Planungsrat für verbindlich erklärt worden (siehe Beschluss 2014/04 der 13. Sitzung). Er wird informell meist als „String.Latin 1.1“ bezeichnet. Die DIN SPEC 91379 als inhaltliche Nachfolgeversion, die nicht nur lateinische, sondern im erweiterten Bereich auch kyrillische und griechische Schriftzeichen enthält, wird aus diesem Grund informell als „String.Latin+ 1.2“ bezeichnet. Dies erklärt auch den Titel dieses informellen Dokuments.

Dieses Dokument „String.Latin+ 1.2“ wird durch die *Fachgruppe String.Latin* erstellt und veröffentlicht. Darin sind Vertreter von Behörden und Organisationseinheiten versammelt, welche die DIN SPEC 91379 in ihren IT Verfahren umsetzen. Sie berät den IT Planungsrat hinsichtlich der Umsetzung der DIN SPEC 91379 in der öffentlichen Verwaltung (siehe Ziffer 1 der Entscheidung 2017/41 des IT-Planungsrats). Sie erreichen die *Fachgruppe String.Latin* unter <String.Latin.Fragen@init.de>.

Die offizielle Fassung der DIN SPEC 91379 inklusive der beigefügten Dateien gemäß Tabelle 2 ist beim Beuth Verlag [<https://www.din.de/de/wdc-beuth:din21:301228458>] kostenfrei erhältlich.

Versionsgeschichte		
Version 1.0	15.1.2020	fgsl
Erste Fassung String.Latin+ 1.2 basierend auf der DIN Spec 91379 vom März 2019		

Inhaltsverzeichnis

Einleitung	ix
1. Anwendungsbereich	1
2. Normative Verweisungen	3
3. Begriffe	5
4. Normative Schriftzeichen	7
4.1. Konformität von IT-Verfahren	7
4.2. Lateinische Buchstaben	7
4.3. Kombinerende diakritische Zeichen	24
4.4. Nicht-Buchstaben N1	25
4.5. Nicht-Buchstaben N2	26
4.6. Nicht-Buchstaben N3	27
4.7. Nicht-Buchstaben N4	28
5. Normative Abbildung lateinischer Buchstaben auf Grundbuchstaben (Suchform)	29
A. Erweiterte Schriftzeichen	51
A.1. Allgemeines	51
A.2. Griechische Buchstaben	51
A.3. Kyrillische Buchstaben	53
A.4. Nicht-Buchstaben E1	55
B. Technische Datentypen	57
B.1. Allgemeines	57
B.2. Datentyp A	59
B.3. Datentyp B	59
B.4. Datentyp C	59
B.5. Datentyp D	59
B.6. Datentyp E	60
B.7. Unicode Normalisierung NFC	60
C. Strukturierte und tabellarische Übersicht aller Zeichen	61
D. Abbildung auf Legacy Codepages	63
E. Häufig gestellte Fragen	65
F. Legacy Zeichensätze	83
G. String.Latin in der Verwaltung	85
Literaturhinweise und Weblinks	91

Abbildungsverzeichnis

B.1. Datentypen der DIN SPEC 91379	58
C.1. Struktur der Datenbasis aller Schriftzeichen dieser DIN SPEC	61

Tabellenverzeichnis

1. Schriftzeichengruppen	xi
2. Angehängte Dateien	xii
3. Normative lateinische Buchstaben	7
4. Normative kombinierende diakritische Zeichen	24
5. Normative Nicht-Buchstaben, erforderlich für alle Namen	25
6. Sonstige normative Nicht-Buchstaben, erforderlich für nicht alle Namen	26
7. Normative Nicht-Buchstaben, erforderlich für Namen juristischer Personen	27
8. Normative Nicht-Buchstaben, für Namen ungeeignet	28
9. Abbildung lateinischer Buchstaben auf Grundbuchstaben analog ICAO	29
A.1. Griechische Buchstaben	51
A.2. Kyrillische Buchstaben	53
A.3. Sonstige erweiterte Nicht-Buchstaben, erforderlich für nicht alle Namen	55
B.1. Übersicht über die empfohlenen Datentypen und zugehörigen Schriftzeichengruppen	58
E.1. Kompatibilitätsmatrix zwischen Schriftarten und Schriftzeichengruppen	78

Einleitung

Ausgangslage

Buchstaben und andere Schriftzeichen (Nicht-Buchstaben) können in IT-Verfahren nur verarbeitet werden, indem sie auf Zahlen abgebildet werden. Es gibt auf der Welt viele verschiedene Alphabete. Für die meisten davon gibt es wiederum viele unterschiedliche Abbildungen der enthaltenen Schriftzeichen auf Zahlen (Zeichenkodierungen oder Encodings). Aus dem Nebeneinander verschiedener Kodierungssysteme für unterschiedliche Alphabete können Interoperabilitätsprobleme resultieren. Der Unicode-Standard bzw. die Norm ISO/IEC 10646 wurde entwickelt, um diese Probleme zu lösen.

Die Forderung nach vollständiger Umsetzung des Unicode-Standards auch auf der Anwendungsebene von IT-Verfahren ist jedoch in der Regel weder sachgerecht noch wirtschaftlich, weil der kulturelle Kontext der Anwender und des intendierten Anwendungsbereichs zu berücksichtigen sind.

IT-Verfahren der öffentlichen Verwaltung Deutschlands sind regelhaft auf die Verarbeitung von Daten in deutscher Sprache und somit in lateinischer Schrift ausgerichtet, denn die Amtssprache ist Deutsch. So legt beispielsweise § 23 Verwaltungsverfahrensgesetz fest: „Werden bei einer Behörde in einer fremden Sprache Anträge gestellt oder Eingaben, Belege, Urkunden oder sonstige Dokumente vorgelegt, soll die Behörde unverzüglich die Vorlage einer Übersetzung verlangen“.

Entsprechend des Übereinkommens über die Angabe von Familiennamen und Vornamen in den Personenstandsbüchern vom 13. September 1973 (BGBl. 1976 II, Seite 1473 f.) sind diese „ohne Übersetzung soweit wie möglich durch Transliteration wiederzugeben. Sind von der Internationalen Normenorganisation (ISO) empfohlene Normen vorhanden, so sind sie anzuwenden“. Sofern die Namen in lateinischen Schriftzeichen vorliegen, „so sind diese Familiennamen und Vornamen buchstabengetreu ohne Änderung oder Übersetzung wiederzugeben“. Diakritische Zeichen sind dementsprechend unverändert wiederzugeben, selbst wenn die deutsche Sprache diese Zeichen nicht kennt.

Diese Vorgaben gelten analog auch für elektronische Register und Datenübermittlungen. Die von der ISO empfohlenen Normen für Transliterationen in die lateinische Schrift sind in den Literaturhinweisen dieses Dokuments aufgeführt.

Standardisierungsbedarf

Aus der Ausgangslage ergibt sich die Notwendigkeit, für IT-Verfahren (insbesondere solche der öffentlichen Verwaltung Deutschlands)

- Einerseits festzulegen, dass alle Schriftzeichen der lateinischen Schrift, die auf hoheitlichen Dokumenten und Urkunden verwendet werden können, vollumfänglich unterstützt werden müssen;
- Andererseits die Möglichkeit zu eröffnen, andere Schriftzeichen von Datenübermittlungen auszuschließen, da diese in den meisten IT-gestützten Prozessen der deutschen Verwaltung nicht akzeptiert werden sollen.

Basierend auf dem Standard Unicode ist somit die Menge der darin enthaltenen, für das Verwaltungshandeln notwendigen und zulässigen Schriftzeichen festzulegen. Diese Teilmenge von Unicode ist anhand der nachfolgend genannten, fachlichen Überlegungen bestimmt worden:

- a. Welche lateinischen Schriftzeichen werden bei der Erfassung und Nacherfassung von Personenstandsdaten für die elektronisch geführten Personenstandsregister festgestellt?
- b. Welche lateinischen Schriftzeichen können als Ergebnis einer Transliteration (das heisst der buchstabengetreuen Übertragung von Wörtern aus einer Schrift in eine andere) aus anderen Schriften auftreten? Dabei müssen die Transliterationsregeln der ISO zugrunde gelegt werden.

- c. Welche lateinischen Schriftzeichen werden in den europäischen Amtssprachen genutzt?
- d. Welche lateinischen Schriftzeichen werden für den Datenaustausch zwischen Sozialversicherungsträgern benötigt?
- e. Welche Schriftzeichen werden für Einträge im Handelsregister benötigt?
- f. Welche Schriftzeichen werden für den EU-weiten Datenaustausch inkl. Island, Liechtenstein, Norwegen und der Schweiz zwischen Behörden benötigt?

Abwägung von Alternativen

Es ist keine Option, die vollständige Unterstützung aller Unicode Blöcke zu fordern, in denen unter anderem lateinische Zeichen enthalten sind, denn:

- Bei der Bestimmung von Zeichenblöcken durch das Unicode-Konsortium spielen die oben genannten, fachlichen Aspekte keine oder nur eine untergeordnete Rolle;
- Es sind auch historische Zeichen ohne Relevanz für das aktuelle Verwaltungshandeln enthalten. Es wäre weder sachgerecht noch wirtschaftlich, deren technische Umsetzung zu fordern;
- Für Transliterationen notwendige kodierte Zeichensequenzen wären nicht explizit im Standard aufgeführt;
- Die Eingabe von ungewöhnlichen Schriftzeichen ist schwieriger und fehleranfälliger, wenn mehr Schriftzeichen unterstützt werden müssen, als gebraucht werden;
- Eine abschließende Liste der zulässigen Schriftzeichen ist eine Voraussetzung für die Implementation robuster Algorithmen und Verfahren, um Personen zu identifizieren und hoheitliche Dokumente und Urkunden zu erstellen. Die in Abschnitt B empfohlenen Datentypen dienen der Unterstützung dieses Ziels.

Zudem entstehen Sicherheitsrisiken, wenn beliebige Kombinationen von Basiszeichen mit diakritischen Zeichen zulässig sind. Dies ermöglicht optisch identische Zeichen, die informationstechnisch unterschiedlich sind. Für eine detaillierte Darstellung wird auf [Unicode TR 36] sowie zur Illustration auf [Confusables] verwiesen. Durch die DIN SPEC lassen sich diese Risiken ausschließen oder zumindest erheblich reduzieren. Denn sie kann in fachlich begründeten Fällen für die Einschränkung genutzt werden, ausschließlich die darin genannten Zeichen zu verarbeiten bzw. zu übermitteln. Die Durchsetzung einer entsprechenden Einschränkung kann sowohl rechtlich als auch technisch erfolgen, beispielsweise durch Nutzung der im Abschnitt B genannten Datentypen.

Die genannten Gründe führen zu dem Ergebnis, dass eine abschließende Auflistung der zu unterstützenden Schriftzeichen als eine echte Teilmenge der in Unicode definierten Zeichen in der DIN SPEC notwendig ist.

Überblick

Im Sinne des Investitionsschutzes berücksichtigt die DIN SPEC den vom IT-Planungsrat und der Innenministerkonferenz derzeit bereits für den gleichen Anwendungsbereich verbindlich vorgegebenen Standard [String Latin 1.1] angemessen.

Die Legacy-Zeichensätze ISO-8859-1, ISO-8859-15 und Windows-1252 waren weitere Grundlagen der Betrachtungen bei der Zusammenstellung der Zeichen der DIN SPEC. Allerdings wurden einige wenige Zeichen aufgrund einer fachlichen Bewertung nicht aufgenommen, die in Anhang F näher erläutert wird.

Die DIN SPEC legt in Abschnitt 4 diejenigen Schriftzeichen fest, die von allen zu der DIN SPEC konformen IT-Verfahren zumindest bei allen für Namen (im weiteren Sinne) bestimmten Datenfeldern unterstützt werden müssen.

Sie empfiehlt in Abschnitt A diejenigen Schriftzeichen, die darüber hinaus von allen zu der DIN SPEC konformen IT-Verfahren unterstützt werden sollten.

Sie legt in Abschnitt 5 fest, wie die in Abschnitt 4 enthaltenen lateinischen Buchstaben auf die lateinischen Grundbuchstaben A bis Z abgebildet werden müssen. Für diesen Zweck werden die Empfehlungen des für maschinenlesbare Reisedokumente einschlägigen Standards [ICAO 9303-3] übernommen und erweitert.

Sie empfiehlt in Abschnitt B Datentypen für Einschränkungen auf bestimmte Teilmengen der in der DIN SPEC genannten Schriftzeichen, die insbesondere bei Schnittstellenvereinbarungen nützlich sein können.

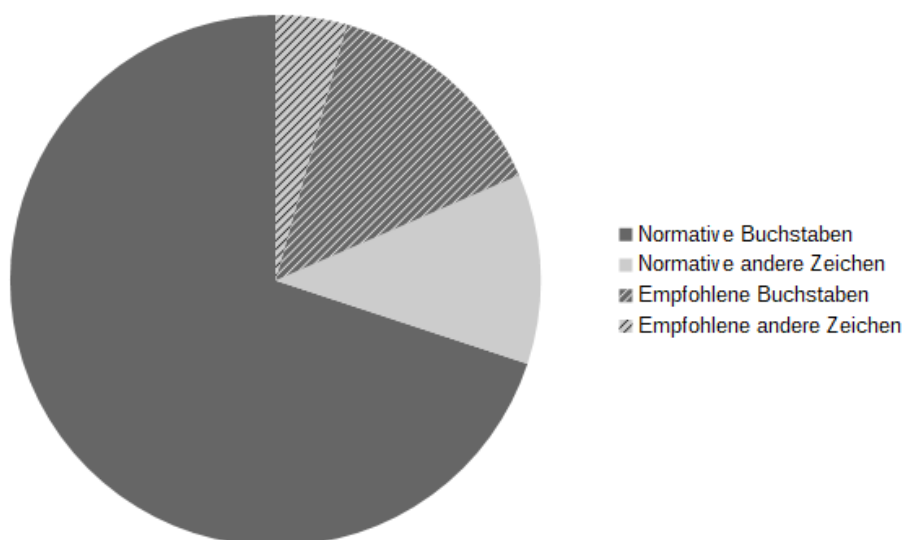
Sie enthält Schriftzeichen in folgenden Schriftzeichengruppen (siehe Tabelle 1):

Tabelle 1. Schriftzeichengruppen


Schriftzeichengruppe	Tabelle	Anzahl Schriftzeichen	davon Zeichensequenzen
Normative lateinische Buchstaben	3	646	147
Normative kombinierende diakritische Zeichen	4	19	0
Normative Nicht-Buchstaben, erforderlich für alle Namen	5	18	0
Sonstige normative Nicht-Buchstaben, erforderlich für nicht alle Namen	6	60	0
Normative Nicht-Buchstaben, erforderlich für Namen juristischer Personen	7	6	0
Normative Nicht-Buchstaben, für Namen ungeeignet	8	4	0
Griechische Buchstaben	A.1	69	0
Kyrillische Buchstaben	A.2	62	0
Sonstige erweiterte Nicht-Buchstaben, erforderlich für nicht alle Namen	A.3	40	0
Summe		924	147

„Lateinische Zeichen“ im Sinne der DIN SPEC sind alle in den Tabellen 3 bis 8 und der Tabelle A.3 enthaltenen Schriftzeichen, also sowohl die Buchstaben als auch die Nicht-Buchstaben.

Von den insgesamt 924 Schriftzeichen sind 753 normativ und 171 empfohlen. Die Verteilung der in der DIN SPEC 91379 enthaltenen Zeichen auf normative und empfohlene Buchstaben sowie normative und empfohlene andere Zeichen ist in dem nachfolgenden Diagramm dargestellt.



Die Schriftzeichengruppen werden in nachfolgenden Abschnitten erläutert und durch tabellarische Aufzählung der darin enthaltenen Schriftzeichen dargestellt. Innerhalb jeder Tabelle der in einer Gruppe enthaltenen Schriftzeichen sind diese nach ihrem Codepoint (bei kodierten Zeichen) bzw. ihren Codepoints (bei kodierten Zeichensequenzen) geordnet. Für jedes kodierte Zeichen wird dessen Codepoint und der vom Unicode Standard zugeordnete Name genannt. Für jede kodierte Zeichensequenz werden deren Codepoints und deren Name genannt. Der angegebene Name der Sequenz ist der vom Unicode Standard festgelegte Name, sofern es sich um eine „named sequence“ gemäß [Unicode 34] handelt. Andernfalls wird der Name unter analoger Anwendung der in Abschnitt 4 [Unicode 34] genannten Regeln zur Bildung des Namens kodierter Zeichensequenzen festgelegt. Alle kodierten Zeichen und Zeichensequenzen werden stets in NFC Normalform angegeben.

Für jedes darstellbare Zeichen wird eine Glyphen angegeben, die an der für die ISO/IEC 10646 gewählten Darstellung orientiert ist. Für jedes nicht darstellbare Zeichen wird ersatzweise die Glyphen „“ (Unicode REPLACEMENT CHARACTER, U+FFFD) verwendet. Die Darstellung der Zeichen durch Glyphen erfolgt mit dem Schriftfont „Calibri Light“ und dient nur der Illustration. Für Implementationen der DIN SPEC muss die Angabe der Codepoints herangezogen werden.

Dieses Dokument wird ergänzt durch die nachfolgend in Tabelle 2 genannten Dateien, die informativ beigelegt werden:

Tabelle 2. Angehängte Dateien

Datei	Beschreibung
din-91379-datatypes.xsd	Eine technische Umsetzung der in Abschnitt 7 beschriebenen Datentypen
latinchars.xml	Eine XML Datei mit allen in dieser DIN SPEC genannten Schriftzeichen und Schriftzeichengruppen
latinchars.xsd	eine XML Schema 1.0 Datei mit der Beschreibung der Struktur der Datei latinchars.xml
2018-11-15_din-spec-91379.xlsx	Eine tabellarische Übersicht über alle in dieser DIN SPEC genannten Schriftzeichen
2018-11-15_legacy-mappings.xlsx	Eine Abbildung der in dieser DIN SPEC enthaltenen Schriftzeichen auf Legacy Codepages

Kapitel 1. Anwendungsbereich

Die DIN SPEC 91379 „Zeichen in Unicode für die elektronische Verarbeitung von Namen und den Datenaustausch in Europa“, und damit auch dieses Dokument „String.Latin+ 1.2“, legen die Zeichen fest, die für Namen im weiteren Sinne benötigt werden. Namen im weiteren Sinne sind Namen natürlicher Personen gemäß der Vorgaben des Personenstandsrechts ebenso wie Namen juristischer Personen, Namen von Produkten, Orten und Straßen sowie Titel von Dokumenten oder Gesetzen. Namen im weiteren Sinne können konkrete Objekte bezeichnen, aber auch virtuelle Konstrukte wie Produktgruppen oder Musikstile.

Die DIN SPEC teilt sich auf in einen normativen Teil (Abschnitte 4 und 5) und einen nicht-normativen Teil.

Der normative Teil der DIN SPEC legt in Abschnitt 4 diejenige Teilmenge der in Unicode enthaltenen Zeichen und Zeichensequenzen fest, die für IT-Verfahren für die elektronische Verarbeitung von Namen (im weiteren Sinne) auf Basis der lateinischen Schrift erforderlich ist. Sie muss deshalb von allen zu dieser Spezifikation konformen IT-Verfahren zumindest bei allen für Namen (im weiteren Sinne) bestimmten Datenfeldern unterstützt werden. Im normativen Teil wird in 5 auch eine gegebenenfalls erforderliche Abbildung der normativen Buchstaben auf die Großbuchstaben A bis Z bestimmt.

Im nicht-normativen Teil der DIN SPEC werden darüber hinaus ergänzende Zeichen aus Unicode festgelegt, die eventuell zur Bildung von Namen juristischer Personen oder von Produkten benötigt werden. Im Kontext des EU-weiten Datenaustauschs ergibt sich ein erweiterter Bedarf an griechischen und kyrillischen Zeichen.

Durch den normativen Teil wird eine Mindestmenge von Schriftzeichen bestimmt. Zur DIN SPEC konforme IT-Verfahren dürfen darüber hinausgehende Schriftzeichen unterstützen. In Schnittstellenvereinbarungen ist es jedoch erforderlich, den jeweiligen Zeichenvorrat abschließend festzulegen. Die DIN SPEC stellt in Abschnitt B Datentypen für fachlich bestimmte Anwendungsfälle zur Verfügung. Dazu werden bestimmte Schriftzeichengruppen zu einer Menge zulässiger Schriftzeichen zusammengestellt. Die informell beigefügte Datei `din-91379-datatypes.xsd` enthält technische Umsetzungen dieser Datentypen, welche ausschließlich die jeweils zugrunde liegenden Schriftzeichen akzeptieren.

Diese DIN SPEC basiert auf dem Unicode-Standard und wendet sich vorrangig an Behörden und Organisationen, die IT-Verfahren betreiben, welche dem behördenübergreifenden Datenaustausch oder dem Datenaustausch mit Bürgern und Wirtschaft dienen. Sie deckt in ihrer Gesamtheit die EU-Amtssprachen und die Amtssprachen Islands, Liechtensteins, Norwegens und der Schweiz sowie die deutschen Minderheitensprachen ab.

Alle in der DIN SPEC genannten Zeichen und Zeichensequenzen liegen in der *Normalization Form C (NFC)* von [Unicode 15] vor.

Die DIN SPEC trifft weder Aussagen zum Umgang mit historischen Zeichen noch zum Umgang mit Fließtext. Die DIN SPEC regelt nicht die Darstellung von Zeichen (Glyphen).

Kapitel 2. Normative Verweisungen

Die folgenden Dokumente werden im Text in solcher Weise in Bezug genommen, dass einige Teile davon oder ihr gesamter Inhalt Anforderungen des vorliegenden Dokuments darstellen. Bei datierten Verweisungen gilt nur die in Bezug genommene Ausgabe. Bei undatierten Verweisungen gilt die letzte Ausgabe des in Bezug genommenen Dokuments (einschließlich aller Änderungen).

ISO/IEC 10646, Information technology — Universal Coded Character Set (UCS) entspricht dem Unicode Standard des Unicode Konsortiums.

Kapitel 3. Begriffe

Die für diese Spezifikation relevanten Definitionen des Unicode-Standards werden unverändert übernommen. Für die Anwendung dieses Dokuments gelten die folgenden Begriffe.

3.1

abstraktes Zeichen

(en: abstract character)

Informationseinheit, die für die Organisation, Kontrolle oder Darstellung von Textdaten verwendet wird

[QUELLE: The Unicode Standard, Version 9.0, 3.4, D.7]

3.2

Codepoint

Wert innerhalb des Wertebereichs der ganzen Zahlen von 0 bis $10FFFF_{16}$

Anmerkung 1 zum Begriff: Der Wert repräsentiert die Position eines Zeichens im Unicode-Standard.

Anmerkung 2 zum Begriff: Für diese DIN SPEC kommen nur Zeichen aus der „Plane 0“ des Unicode-Standards zum Einsatz, sodass der Wert eines Codepoints $FFFF_{16}$ nicht überschreiten kann.

[QUELLE: The Unicode Standard, Version 9.0, 3.4, D.10 in Verbindung mit D.9]

3.3

kodiertes Zeichen

(en: encoded character, coded character)

Zeichen

(en: character)

Verbindung (oder Abbildung) zwischen einem abstrakten Zeichen und einem Codepoint

Anmerkung 1 zum Begriff: Der Begriff „Zeichen“ ist allein stehend als Abkürzung für den Begriff „kodiertes Zeichen“ zu verstehen. Siehe hierfür die Erläuterung in Unicode 9.0, 3.4, zwischen den Definitionen D.12 und D.13.

[QUELLE: The Unicode Standard, Version 9.0, 3.4, D.11]

3.4

kodierte Zeichensequenz

(en: coded character sequence)

Zeichensequenz

(en: character sequence)

geordnete Sequenz eines oder mehrerer Codepoints

Anmerkung 1 zum Begriff: Der Begriff „Zeichensequenz“ steht als Abkürzung für den Begriff „kodierte Zeichensequenz“. Siehe hierfür die Erläuterung in [Unicode 9], 3.4, zwischen den Definitionen D.12 und D.13

[QUELLE: The Unicode Standard, Version 9.0, 3.4, D.12]

3.5

Schriftzeichen

kleinste im Schreibfluss aufeinanderfolgende Einheiten einer Schrift, z. B. des lateinischen oder kyrillischen Alphabets, die sprachunabhängig sind, da sie für verschiedene Sprachen genutzt werden können, z. B. das lateinische Alphabet für Deutsch und Französisch

Anmerkung 1 zum Begriff: Zu den Schriftzeichen gehören sowohl kodierte Zeichen als auch kodierte Zeichensequenzen, jedoch keine einzelnen kombinierenden diakritischen Zeichen außerhalb von Zeichensequenzen.

Anmerkung 2 zum Begriff: Im Rahmen dieser DIN SPEC werden die Schriftzeichen unterschieden in Buchstaben und Nicht-Buchstaben (z. B. Interpunktion, Ziffern und Leerzeichen).

Anmerkung 3 zum Begriff: Die Menge der Schriftzeichen, deren Unterstützung durch IT-Verfahren für Zwecke der elektronischen Verarbeitung von Namen und den Datenaustausch in Europa vorgeschrieben bzw. empfohlen wird, ist eine echte Teilmenge der im Unicode-Standard insgesamt enthaltenen Schriftzeichen.

Anmerkung 4 zum Begriff: In dieser DIN SPEC wird der Begriff „Schriftzeichen“ weitgehend synonym zum Begriff „Graphem“ verwendet. Es werden aber auch „Plurigraphie“ mit eingeschlossen.

3.6

Schriftzeichengruppe

Zusammenfassung von Schriftzeichen unter fachlichen Aspekten

Anmerkung 1 zum Begriff: Jedes der in dieser Spezifikation genannten Schriftzeichen ist genau einer Schriftzeichengruppe zugeordnet. Jeder Schriftzeichengruppe ist mindestens ein Schriftzeichen zugeordnet. Jede Schriftzeichengruppe hat einen innerhalb dieser Spezifikation eindeutigen Namen, z. B. „Lateinische Buchstaben“ oder „Nicht-Buchstaben N1“.

Anmerkung 2 zum Begriff: In Tabelle 1 sind alle Schriftzeichengruppen genannt.

Kapitel 4. Normative Schriftzeichen

4.1. Konformität von IT-Verfahren

Jedes konforme IT-Verfahren muss die in Tabelle 3 bis Tabelle 8 genannten Schriftzeichen bei allen für Namen (im weiteren Sinne) bestimmten Datenfeldern in vollem Umfang verarbeiten können. Dies umfasst insbesondere die Erfassung, Speicherung, Übermittlung, Anzeige und den Ausdruck aller Schriftzeichen.

In Abschnitt E werden Formulierungsvorschläge für die Verwendung in Ausschreibungen zur Beschaffung konformer IT Verfahren vorgeschlagen.

4.2. Lateinische Buchstaben

Die in Tabelle 3 genannten Schriftzeichen für lateinische Buchstaben sind für die Darstellung von Namen im weiteren Sinne zwingend erforderlich.

Tabelle 3. Normative lateinische Buchstaben

Codepoints	Name	Glyphe
U+0041	LATIN CAPITAL LETTER A	A
<U+0041,U+030B>	LATIN CAPITAL LETTER A WITH COMBINING DOUBLE ACUTE ACCENT	Ǻ
U+0042	LATIN CAPITAL LETTER B	B
U+0043	LATIN CAPITAL LETTER C	C
<U+0043,U+0300>	LATIN CAPITAL LETTER C WITH COMBINING GRAVE ACCENT	Ĉ
<U+0043,U+0304>	LATIN CAPITAL LETTER C WITH COMBINING MACRON	Ċ
<U+0043,U+0306>	LATIN CAPITAL LETTER C WITH COMBINING BREVE	Č
<U+0043,U+0308>	LATIN CAPITAL LETTER C WITH COMBINING DIAERESIS	Č̈
<U+0043,U+0315>	LATIN CAPITAL LETTER C WITH COMBINING COMMA ABOVE RIGHT	Ć
<U+0043,U+0323>	LATIN CAPITAL LETTER C WITH COMBINING DOT BELOW	Ḙ
<U+0043,U+0326>	LATIN CAPITAL LETTER C WITH COMBINING COMMA BELOW	Ḛ
<U+0043,U+0328,U+0306>	LATIN CAPITAL LETTER C WITH COMBINING OGONEK AND COMBINING BREVE	Ć̇
U+0044	LATIN CAPITAL LETTER D	D
<U+0044,U+0302>	LATIN CAPITAL LETTER D WITH COMBINING CIRCUMFLEX ACCENT	Ď
U+0045	LATIN CAPITAL LETTER E	E
U+0046	LATIN CAPITAL LETTER F	F
<U+0046,U+0300>	LATIN CAPITAL LETTER F WITH COMBINING GRAVE ACCENT	Ĕ
<U+0046,U+0304>	LATIN CAPITAL LETTER F WITH COMBINING MACRON	Ĳ
U+0047	LATIN CAPITAL LETTER G	G
<U+0047,U+0300>	LATIN CAPITAL LETTER G WITH COMBINING GRAVE ACCENT	Ĝ
U+0048	LATIN CAPITAL LETTER H	H
<U+0048,U+0304>	LATIN CAPITAL LETTER H WITH COMBINING MACRON	Ĥ
<U+0048,U+0326>	LATIN CAPITAL LETTER H WITH COMBINING COMMA BELOW	Ḧ

Codepoints	Name	Glyphe
<U+0048,U+0331>	LATIN CAPITAL LETTER H WITH COMBINING MACRON BELOW	Ĥ
U+0049	LATIN CAPITAL LETTER I	I
U+004A	LATIN CAPITAL LETTER J	J
<U+004A,U+0301>	LATIN CAPITAL LETTER J WITH COMBINING ACUTE ACCENT	Ĵ
<U+004A,U+030C>	LATIN CAPITAL LETTER J WITH COMBINING CARON	Ĵ̇
U+004B	LATIN CAPITAL LETTER K	K
<U+004B,U+0300>	LATIN CAPITAL LETTER K WITH COMBINING GRAVE ACCENT	Ķ
<U+004B,U+0302>	LATIN CAPITAL LETTER K WITH COMBINING CIRCUMFLEX ACCENT	Ĳ
<U+004B,U+0304>	LATIN CAPITAL LETTER K WITH COMBINING MACRON	Ĳ̄
<U+004B,U+0307>	LATIN CAPITAL LETTER K WITH COMBINING DOT ABOVE	Ĳ̣
<U+004B,U+0315>	LATIN CAPITAL LETTER K WITH COMBINING COMMA ABOVE RIGHT	Ķ́
<U+004B,U+031B>	LATIN CAPITAL LETTER K WITH COMBINING HORN	Რ
<U+004B,U+0326>	LATIN CAPITAL LETTER K WITH COMBINING COMMA BELOW	Რ̸
<U+004B,U+035F,U+0048>	LATIN CAPITAL LETTER K WITH COMBINING DOUBLE MACRON BELOW AND LATIN CAPITAL LETTER H	Რ̸H
<U+004B,U+035F,U+0068>	LATIN CAPITAL LETTER K WITH COMBINING DOUBLE MACRON BELOW AND LATIN SMALL LETTER H	Რ̸h
U+004C	LATIN CAPITAL LETTER L	L
<U+004C,U+0302>	LATIN CAPITAL LETTER L WITH COMBINING CIRCUMFLEX ACCENT	Ĺ
<U+004C,U+0325>	LATIN CAPITAL LETTER L WITH COMBINING RING BELOW	Რ̣
<U+004C,U+0325,U+0304>	LATIN CAPITAL LETTER L WITH COMBINING RING BELOW AND COMBINING MACRON	Რ̣̄
<U+004C,U+0326>	LATIN CAPITAL LETTER L WITH COMBINING COMMA BELOW	Რ̸
U+004D	LATIN CAPITAL LETTER M	M
<U+004D,U+0300>	LATIN CAPITAL LETTER M WITH COMBINING GRAVE ACCENT	Ṁ
<U+004D,U+0302>	LATIN CAPITAL LETTER M WITH COMBINING CIRCUMFLEX ACCENT	Ṃ
<U+004D,U+0306>	LATIN CAPITAL LETTER M WITH COMBINING BREVE	Ṅ
<U+004D,U+0310>	LATIN CAPITAL LETTER M WITH COMBINING CANDRABINDU	Ṃ̣
U+004E	LATIN CAPITAL LETTER N	N
<U+004E,U+0302>	LATIN CAPITAL LETTER N WITH COMBINING CIRCUMFLEX ACCENT	Ñ
<U+004E,U+0304>	LATIN CAPITAL LETTER N WITH COMBINING MACRON	Ñ̄
<U+004E,U+0306>	LATIN CAPITAL LETTER N WITH COMBINING BREVE	Ñ̆
<U+004E,U+0326>	LATIN CAPITAL LETTER N WITH COMBINING COMMA BELOW	Ṇ̸
U+004F	LATIN CAPITAL LETTER O	O
U+0050	LATIN CAPITAL LETTER P	P
<U+0050,U+0300>	LATIN CAPITAL LETTER P WITH COMBINING GRAVE ACCENT	Ṗ
<U+0050,U+0304>	LATIN CAPITAL LETTER P WITH COMBINING MACRON	Ṗ̄
<U+0050,U+0315>	LATIN CAPITAL LETTER P WITH COMBINING COMMA ABOVE RIGHT	Ṗ́
<U+0050,U+0323>	LATIN CAPITAL LETTER P WITH COMBINING DOT BELOW	Ṗ̣
U+0051	LATIN CAPITAL LETTER Q	Q

Codepoints	Name	Glyphe
U+0052	LATIN CAPITAL LETTER R	R
<U+0052,U+0306>	LATIN CAPITAL LETTER R WITH COMBINING BREVE	Ř
<U+0052,U+0325>	LATIN CAPITAL LETTER R WITH COMBINING RING BELOW	Ŕ
<U+0052,U+0325,U+0304>	LATIN CAPITAL LETTER R WITH COMBINING RING BELOW AND COMBINING MACRON	Ŗ
U+0053	LATIN CAPITAL LETTER S	S
<U+0053,U+0300>	LATIN CAPITAL LETTER S WITH COMBINING GRAVE ACCENT	Š
<U+0053,U+0304>	LATIN CAPITAL LETTER S WITH COMBINING MACRON	Ŝ
<U+0053,U+031B,U+0304>	LATIN CAPITAL LETTER S WITH COMBINING HORN AND COMBINING MACRON	Ś
<U+0053,U+0331>	LATIN CAPITAL LETTER S WITH COMBINING MACRON BELOW	Ṣ
U+0054	LATIN CAPITAL LETTER T	T
<U+0054,U+0300>	LATIN CAPITAL LETTER T WITH COMBINING GRAVE ACCENT	Ť
<U+0054,U+0304>	LATIN CAPITAL LETTER T WITH COMBINING MACRON	Ṫ
<U+0054,U+0308>	LATIN CAPITAL LETTER T WITH COMBINING DIAERESIS	Ŧ
<U+0054,U+0315>	LATIN CAPITAL LETTER T WITH COMBINING COMMA ABOVE RIGHT	Ṭ
<U+0054,U+031B>	LATIN CAPITAL LETTER T WITH COMBINING HORN	Ṣ
U+0055	LATIN CAPITAL LETTER U	U
<U+0055,U+0307>	LATIN CAPITAL LETTER U WITH COMBINING DOT ABOVE	Ů
U+0056	LATIN CAPITAL LETTER V	V
U+0057	LATIN CAPITAL LETTER W	W
U+0058	LATIN CAPITAL LETTER X	X
U+0059	LATIN CAPITAL LETTER Y	Y
U+005A	LATIN CAPITAL LETTER Z	Z
<U+005A,U+0300>	LATIN CAPITAL LETTER Z WITH COMBINING GRAVE ACCENT	Ž
<U+005A,U+0304>	LATIN CAPITAL LETTER Z WITH COMBINING MACRON	Ẑ
<U+005A,U+0306>	LATIN CAPITAL LETTER Z WITH COMBINING BREVE	Ẓ
<U+005A,U+0308>	LATIN CAPITAL LETTER Z WITH COMBINING DIAERESIS	Ẕ
<U+005A,U+0327>	LATIN CAPITAL LETTER Z WITH COMBINING CEDILLA	Ẑ
U+0061	LATIN SMALL LETTER A	a
<U+0061,U+030B>	LATIN SMALL LETTER A WITH COMBINING DOUBLE ACUTE ACCENT	Ǻ
U+0062	LATIN SMALL LETTER B	b
U+0063	LATIN SMALL LETTER C	c
<U+0063,U+0300>	LATIN SMALL LETTER C WITH COMBINING GRAVE ACCENT	ç
<U+0063,U+0304>	LATIN SMALL LETTER C WITH COMBINING MACRON	ċ
<U+0063,U+0306>	LATIN SMALL LETTER C WITH COMBINING BREVE	ċ
<U+0063,U+0308>	LATIN SMALL LETTER C WITH COMBINING DIAERESIS	ċ
<U+0063,U+0315>	LATIN SMALL LETTER C WITH COMBINING COMMA ABOVE RIGHT	c´
<U+0063,U+0323>	LATIN SMALL LETTER C WITH COMBINING DOT BELOW	ç
<U+0063,U+0326>	LATIN SMALL LETTER C WITH COMBINING COMMA BELOW	ç

Codepoints	Name	Glyphe
<U+0063,U+0328,U+0306>	LATIN SMALL LETTER C WITH COMBINING OGONEK AND COMBINING BREVE	č
U+0064	LATIN SMALL LETTER D	d
<U+0064,U+0302>	LATIN SMALL LETTER D WITH COMBINING CIRCUMFLEX ACCENT	â
U+0065	LATIN SMALL LETTER E	e
U+0066	LATIN SMALL LETTER F	f
<U+0066,U+0300>	LATIN SMALL LETTER F WITH COMBINING GRAVE ACCENT	ƒ
<U+0066,U+0304>	LATIN SMALL LETTER F WITH COMBINING MACRON	ƒ̄
U+0067	LATIN SMALL LETTER G	g
<U+0067,U+0300>	LATIN SMALL LETTER G WITH COMBINING GRAVE ACCENT	ğ
U+0068	LATIN SMALL LETTER H	h
<U+0068,U+0304>	LATIN SMALL LETTER H WITH COMBINING MACRON	h̄
<U+0068,U+0326>	LATIN SMALL LETTER H WITH COMBINING COMMA BELOW	h̸
U+0069	LATIN SMALL LETTER I	i
U+006A	LATIN SMALL LETTER J	j
<U+006A,U+0301>	LATIN SMALL LETTER J WITH COMBINING ACUTE ACCENT	í
U+006B	LATIN SMALL LETTER K	k
<U+006B,U+0300>	LATIN SMALL LETTER K WITH COMBINING GRAVE ACCENT	ķ
<U+006B,U+0302>	LATIN SMALL LETTER K WITH COMBINING CIRCUMFLEX ACCENT	ķ̂
<U+006B,U+0304>	LATIN SMALL LETTER K WITH COMBINING MACRON	ķ̄
<U+006B,U+0307>	LATIN SMALL LETTER K WITH COMBINING DOT ABOVE	ķ̣
<U+006B,U+0315>	LATIN SMALL LETTER K WITH COMBINING COMMA ABOVE RIGHT	ķ̸ʹ
<U+006B,U+031B>	LATIN SMALL LETTER K WITH COMBINING HORN	ķ̣̈́
<U+006B,U+0326>	LATIN SMALL LETTER K WITH COMBINING COMMA BELOW	ķ̸
<U+006B,U+035F,U+0068>	LATIN SMALL LETTER K WITH COMBINING DOUBLE MACRON BELOW AND LATIN SMALL LETTER H	kh̶
U+006C	LATIN SMALL LETTER L	l
<U+006C,U+0302>	LATIN SMALL LETTER L WITH COMBINING CIRCUMFLEX ACCENT	l̂
<U+006C,U+0325>	LATIN SMALL LETTER L WITH COMBINING RING BELOW	ł̣
<U+006C,U+0325,U+0304>	LATIN SMALL LETTER L WITH COMBINING RING BELOW AND COMBINING MACRON	ł̣̄
<U+006C,U+0326>	LATIN SMALL LETTER L WITH COMBINING COMMA BELOW	ł̸
U+006D	LATIN SMALL LETTER M	m
<U+006D,U+0300>	LATIN SMALL LETTER M WITH COMBINING GRAVE ACCENT	ṁ
<U+006D,U+0302>	LATIN SMALL LETTER M WITH COMBINING CIRCUMFLEX ACCENT	Ṃ
<U+006D,U+0306>	LATIN SMALL LETTER M WITH COMBINING BREVE	ṃ
<U+006D,U+0310>	LATIN SMALL LETTER M WITH COMBINING CANDRABINDU	Ṅ
U+006E	LATIN SMALL LETTER N	n
<U+006E,U+0302>	LATIN SMALL LETTER N WITH COMBINING CIRCUMFLEX ACCENT	ñ
<U+006E,U+0304>	LATIN SMALL LETTER N WITH COMBINING MACRON	n̄

Codepoints	Name	Glyphe
<U+006E,U+0306>	LATIN SMALL LETTER N WITH COMBINING BREVE	ñ
<U+006E,U+0326>	LATIN SMALL LETTER N WITH COMBINING COMMA BELOW	ŋ
U+006F	LATIN SMALL LETTER O	o
U+0070	LATIN SMALL LETTER P	p
<U+0070,U+0300>	LATIN SMALL LETTER P WITH COMBINING GRAVE ACCENT	ò
<U+0070,U+0304>	LATIN SMALL LETTER P WITH COMBINING MACRON	ō
<U+0070,U+0315>	LATIN SMALL LETTER P WITH COMBINING COMMA ABOVE RIGHT	Ɔ
<U+0070,U+0323>	LATIN SMALL LETTER P WITH COMBINING DOT BELOW	Ṗ
U+0071	LATIN SMALL LETTER Q	q
U+0072	LATIN SMALL LETTER R	r
<U+0072,U+0306>	LATIN SMALL LETTER R WITH COMBINING BREVE	ř
<U+0072,U+0325>	LATIN SMALL LETTER R WITH COMBINING RING BELOW	ŕ
<U+0072,U+0325,U+0304>	LATIN SMALL LETTER R WITH COMBINING RING BELOW AND COMBINING MACRON	ṙ
U+0073	LATIN SMALL LETTER S	s
<U+0073,U+0300>	LATIN SMALL LETTER S WITH COMBINING GRAVE ACCENT	ș
<U+0073,U+0304>	LATIN SMALL LETTER S WITH COMBINING MACRON	ṡ
<U+0073,U+031B,U+0304>	LATIN SMALL LETTER S WITH COMBINING HORN AND COMBINING MACRON	Ṣ
<U+0073,U+0331>	LATIN SMALL LETTER S WITH COMBINING MACRON BELOW	ṣ
U+0074	LATIN SMALL LETTER T	t
<U+0074,U+0300>	LATIN SMALL LETTER T WITH COMBINING GRAVE ACCENT	ť
<U+0074,U+0304>	LATIN SMALL LETTER T WITH COMBINING MACRON	ṭ
<U+0074,U+0315>	LATIN SMALL LETTER T WITH COMBINING COMMA ABOVE RIGHT	Ƨ
<U+0074,U+031B>	LATIN SMALL LETTER T WITH COMBINING HORN	Ṫ
U+0075	LATIN SMALL LETTER U	u
<U+0075,U+0307>	LATIN SMALL LETTER U WITH COMBINING DOT ABOVE	û
U+0076	LATIN SMALL LETTER V	v
U+0077	LATIN SMALL LETTER W	w
U+0078	LATIN SMALL LETTER X	x
U+0079	LATIN SMALL LETTER Y	y
U+007A	LATIN SMALL LETTER Z	z
<U+007A,U+0300>	LATIN SMALL LETTER Z WITH COMBINING GRAVE ACCENT	ž
<U+007A,U+0304>	LATIN SMALL LETTER Z WITH COMBINING MACRON	ẓ
<U+007A,U+0306>	LATIN SMALL LETTER Z WITH COMBINING BREVE	ẏ
<U+007A,U+0308>	LATIN SMALL LETTER Z WITH COMBINING DIAERESIS	Ẓ
<U+007A,U+0327>	LATIN SMALL LETTER Z WITH COMBINING CEDILLA	Ẕ
U+00C0	LATIN CAPITAL LETTER A WITH GRAVE	À
U+00C1	LATIN CAPITAL LETTER A WITH ACUTE	Á
U+00C2	LATIN CAPITAL LETTER A WITH CIRCUMFLEX	Â

Codepoints	Name	Glyphe
U+00C3	LATIN CAPITAL LETTER A WITH TILDE	Ã
U+00C4	LATIN CAPITAL LETTER A WITH DIAERESIS	Ä
U+00C5	LATIN CAPITAL LETTER A WITH RING ABOVE	Å
U+00C6	LATIN CAPITAL LETTER AE	Æ
U+00C7	LATIN CAPITAL LETTER C WITH CEDILLA	Ç
<U+00C7,U+0306>	LATIN CAPITAL LETTER C WITH CEDILLA WITH COMBINING BREVE	Č
U+00C8	LATIN CAPITAL LETTER E WITH GRAVE	È
U+00C9	LATIN CAPITAL LETTER E WITH ACUTE	É
U+00CA	LATIN CAPITAL LETTER E WITH CIRCUMFLEX	Ê
U+00CB	LATIN CAPITAL LETTER E WITH DIAERESIS	Ë
U+00CC	LATIN CAPITAL LETTER I WITH GRAVE	Ì
U+00CD	LATIN CAPITAL LETTER I WITH ACUTE	Í
U+00CE	LATIN CAPITAL LETTER I WITH CIRCUMFLEX	Î
U+00CF	LATIN CAPITAL LETTER I WITH DIAERESIS	Ï
U+00D0	LATIN CAPITAL LETTER ETH	Ð
U+00D1	LATIN CAPITAL LETTER N WITH TILDE	Ñ
U+00D2	LATIN CAPITAL LETTER O WITH GRAVE	Ò
U+00D3	LATIN CAPITAL LETTER O WITH ACUTE	Ó
U+00D4	LATIN CAPITAL LETTER O WITH CIRCUMFLEX	Ô
U+00D5	LATIN CAPITAL LETTER O WITH TILDE	Õ
U+00D6	LATIN CAPITAL LETTER O WITH DIAERESIS	Ö
U+00D8	LATIN CAPITAL LETTER O WITH STROKE	Ø
U+00D9	LATIN CAPITAL LETTER U WITH GRAVE	Ù
U+00DA	LATIN CAPITAL LETTER U WITH ACUTE	Ú
U+00DB	LATIN CAPITAL LETTER U WITH CIRCUMFLEX	Û
<U+00DB,U+0304>	LATIN CAPITAL LETTER U WITH CIRCUMFLEX WITH COMBINING MACRON	Ů
U+00DC	LATIN CAPITAL LETTER U WITH DIAERESIS	Ü
U+00DD	LATIN CAPITAL LETTER Y WITH ACUTE	Ý
U+00DE	LATIN CAPITAL LETTER THORN	Þ
U+00DF	LATIN SMALL LETTER SHARP S	ß
U+00E0	LATIN SMALL LETTER A WITH GRAVE	à
U+00E1	LATIN SMALL LETTER A WITH ACUTE	á
U+00E2	LATIN SMALL LETTER A WITH CIRCUMFLEX	â
U+00E3	LATIN SMALL LETTER A WITH TILDE	ã
U+00E4	LATIN SMALL LETTER A WITH DIAERESIS	ä
U+00E5	LATIN SMALL LETTER A WITH RING ABOVE	å
U+00E6	LATIN SMALL LETTER AE	æ
U+00E7	LATIN SMALL LETTER C WITH CEDILLA	ç
<U+00E7,U+0306>	LATIN SMALL LETTER C WITH CEDILLA WITH COMBINING BREVE	č

Codepoints	Name	Glyphe
U+00E8	LATIN SMALL LETTER E WITH GRAVE	è
U+00E9	LATIN SMALL LETTER E WITH ACUTE	é
U+00EA	LATIN SMALL LETTER E WITH CIRCUMFLEX	ê
U+00EB	LATIN SMALL LETTER E WITH DIAERESIS	ë
U+00EC	LATIN SMALL LETTER I WITH GRAVE	ì
U+00ED	LATIN SMALL LETTER I WITH ACUTE	í
U+00EE	LATIN SMALL LETTER I WITH CIRCUMFLEX	î
U+00EF	LATIN SMALL LETTER I WITH DIAERESIS	ï
U+00F0	LATIN SMALL LETTER ETH	ð
U+00F1	LATIN SMALL LETTER N WITH TILDE	ñ
U+00F2	LATIN SMALL LETTER O WITH GRAVE	ò
U+00F3	LATIN SMALL LETTER O WITH ACUTE	ó
U+00F4	LATIN SMALL LETTER O WITH CIRCUMFLEX	ô
U+00F5	LATIN SMALL LETTER O WITH TILDE	õ
U+00F6	LATIN SMALL LETTER O WITH DIAERESIS	ö
U+00F8	LATIN SMALL LETTER O WITH STROKE	ø
U+00F9	LATIN SMALL LETTER U WITH GRAVE	ù
U+00FA	LATIN SMALL LETTER U WITH ACUTE	ú
U+00FB	LATIN SMALL LETTER U WITH CIRCUMFLEX	û
<U+00FB,U+0304>	LATIN SMALL LETTER U WITH CIRCUMFLEX WITH COMBINING MACRON	ũ
U+00FC	LATIN SMALL LETTER U WITH DIAERESIS	ü
U+00FD	LATIN SMALL LETTER Y WITH ACUTE	ý
U+00FE	LATIN SMALL LETTER THORN	þ
U+00FF	LATIN SMALL LETTER Y WITH DIAERESIS	ÿ
<U+00FF,U+0301>	LATIN SMALL LETTER Y WITH DIAERESIS WITH COMBINING ACUTE ACCENT	ÿ́
U+0100	LATIN CAPITAL LETTER A WITH MACRON	Ā
U+0101	LATIN SMALL LETTER A WITH MACRON	ā
U+0102	LATIN CAPITAL LETTER A WITH BREVE	Ă
U+0103	LATIN SMALL LETTER A WITH BREVE	ă
U+0104	LATIN CAPITAL LETTER A WITH OGONEK	Ą
U+0105	LATIN SMALL LETTER A WITH OGONEK	ą
U+0106	LATIN CAPITAL LETTER C WITH ACUTE	Ć
U+0107	LATIN SMALL LETTER C WITH ACUTE	ć
U+0108	LATIN CAPITAL LETTER C WITH CIRCUMFLEX	Ĉ
U+0109	LATIN SMALL LETTER C WITH CIRCUMFLEX	ĉ
U+010A	LATIN CAPITAL LETTER C WITH DOT ABOVE	Č
U+010B	LATIN SMALL LETTER C WITH DOT ABOVE	č
U+010C	LATIN CAPITAL LETTER C WITH CARON	Č

Codepoints	Name	Glyphe
<U+010C,U+0315>	LATIN CAPITAL LETTER C WITH CARON WITH COMBINING COMMA ABOVE RIGHT	Č̣
<U+010C,U+0323>	LATIN CAPITAL LETTER C WITH CARON WITH COMBINING DOT BELOW	Č̣̣
U+010D	LATIN SMALL LETTER C WITH CARON	č
<U+010D,U+0315>	LATIN SMALL LETTER C WITH CARON WITH COMBINING COMMA ABOVE RIGHT	č̣
<U+010D,U+0323>	LATIN SMALL LETTER C WITH CARON WITH COMBINING DOT BELOW	č̣̣
U+010E	LATIN CAPITAL LETTER D WITH CARON	Ď
U+010F	LATIN SMALL LETTER D WITH CARON	ď
U+0110	LATIN CAPITAL LETTER D WITH STROKE	Đ
U+0111	LATIN SMALL LETTER D WITH STROKE	đ
U+0112	LATIN CAPITAL LETTER E WITH MACRON	Ē
U+0113	LATIN SMALL LETTER E WITH MACRON	ē
U+0114	LATIN CAPITAL LETTER E WITH BREVE	Ė
U+0115	LATIN SMALL LETTER E WITH BREVE	ė
U+0116	LATIN CAPITAL LETTER E WITH DOT ABOVE	Ê
U+0117	LATIN SMALL LETTER E WITH DOT ABOVE	ê
U+0118	LATIN CAPITAL LETTER E WITH OGONEK	Ę
U+0119	LATIN SMALL LETTER E WITH OGONEK	ę
U+011A	LATIN CAPITAL LETTER E WITH CARON	Ě
U+011B	LATIN SMALL LETTER E WITH CARON	ě
U+011C	LATIN CAPITAL LETTER G WITH CIRCUMFLEX	Ĝ
U+011D	LATIN SMALL LETTER G WITH CIRCUMFLEX	ĝ
U+011E	LATIN CAPITAL LETTER G WITH BREVE	Ğ
U+011F	LATIN SMALL LETTER G WITH BREVE	ğ
U+0120	LATIN CAPITAL LETTER G WITH DOT ABOVE	Ġ
U+0121	LATIN SMALL LETTER G WITH DOT ABOVE	ġ
U+0122	LATIN CAPITAL LETTER G WITH CEDILLA	Ģ
U+0123	LATIN SMALL LETTER G WITH CEDILLA	ģ
U+0124	LATIN CAPITAL LETTER H WITH CIRCUMFLEX	Ĥ
U+0125	LATIN SMALL LETTER H WITH CIRCUMFLEX	ĥ
U+0126	LATIN CAPITAL LETTER H WITH STROKE	Ħ
U+0127	LATIN SMALL LETTER H WITH STROKE	ħ
U+0128	LATIN CAPITAL LETTER I WITH TILDE	Ĩ
U+0129	LATIN SMALL LETTER I WITH TILDE	ĩ
U+012A	LATIN CAPITAL LETTER I WITH MACRON	Ī
<U+012A,U+0301>	LATIN CAPITAL LETTER I WITH MACRON WITH COMBINING ACUTE ACCENT	Ī́
U+012B	LATIN SMALL LETTER I WITH MACRON	ī

Codepoints	Name	Glyphe
<U+012B,U+0301>	LATIN SMALL LETTER I WITH MACRON WITH COMBINING ACUTE ACCENT	í
U+012C	LATIN CAPITAL LETTER I WITH BREVE	İ
U+012D	LATIN SMALL LETTER I WITH BREVE	ı
U+012E	LATIN CAPITAL LETTER I WITH OGONEK	Į
U+012F	LATIN SMALL LETTER I WITH OGONEK	į
U+0130	LATIN CAPITAL LETTER I WITH DOT ABOVE	İ
U+0131	LATIN SMALL LETTER DOTLESS I	ı
U+0132	LATIN CAPITAL LIGATURE IJ	IJ
U+0133	LATIN SMALL LIGATURE IJ	ij
U+0134	LATIN CAPITAL LETTER J WITH CIRCUMFLEX	Ĵ
U+0135	LATIN SMALL LETTER J WITH CIRCUMFLEX	ĵ
U+0136	LATIN CAPITAL LETTER K WITH CEDILLA	Ķ
U+0137	LATIN SMALL LETTER K WITH CEDILLA	ķ
U+0138	LATIN SMALL LETTER KRA	κ
U+0139	LATIN CAPITAL LETTER L WITH ACUTE	Ł
U+013A	LATIN SMALL LETTER L WITH ACUTE	ł
U+013B	LATIN CAPITAL LETTER L WITH CEDILLA	Ľ
U+013C	LATIN SMALL LETTER L WITH CEDILLA	ĺ
U+013D	LATIN CAPITAL LETTER L WITH CARON	Ĺ
U+013E	LATIN SMALL LETTER L WITH CARON	ľ
U+013F	LATIN CAPITAL LETTER L WITH MIDDLE DOT	Ľ
U+0140	LATIN SMALL LETTER L WITH MIDDLE DOT	ḷ
U+0141	LATIN CAPITAL LETTER L WITH STROKE	Ł
U+0142	LATIN SMALL LETTER L WITH STROKE	ł
U+0143	LATIN CAPITAL LETTER N WITH ACUTE	Ń
U+0144	LATIN SMALL LETTER N WITH ACUTE	ń
U+0145	LATIN CAPITAL LETTER N WITH CEDILLA	Ñ
U+0146	LATIN SMALL LETTER N WITH CEDILLA	ñ
U+0147	LATIN CAPITAL LETTER N WITH CARON	Ň
U+0148	LATIN SMALL LETTER N WITH CARON	ň
U+0149	LATIN SMALL LETTER N PRECEDED BY APOSTROPHE	’n
U+014A	LATIN CAPITAL LETTER ENG	Ŋ
U+014B	LATIN SMALL LETTER ENG	ŋ
U+014C	LATIN CAPITAL LETTER O WITH MACRON	Ō
U+014D	LATIN SMALL LETTER O WITH MACRON	ō
U+014E	LATIN CAPITAL LETTER O WITH BREVE	Ȫ
U+014F	LATIN SMALL LETTER O WITH BREVE	ȫ
U+0150	LATIN CAPITAL LETTER O WITH DOUBLE ACUTE	Ő
U+0151	LATIN SMALL LETTER O WITH DOUBLE ACUTE	ő

Codepoints	Name	Glyphe
U+0152	LATIN CAPITAL LIGATURE OE	Œ
U+0153	LATIN SMALL LIGATURE OE	œ
U+0154	LATIN CAPITAL LETTER R WITH ACUTE	Ŕ
U+0155	LATIN SMALL LETTER R WITH ACUTE	ŕ
U+0156	LATIN CAPITAL LETTER R WITH CEDILLA	Ŗ
U+0157	LATIN SMALL LETTER R WITH CEDILLA	ŗ
U+0158	LATIN CAPITAL LETTER R WITH CARON	Ř
U+0159	LATIN SMALL LETTER R WITH CARON	ř
U+015A	LATIN CAPITAL LETTER S WITH ACUTE	Ŝ
U+015B	LATIN SMALL LETTER S WITH ACUTE	ŝ
U+015C	LATIN CAPITAL LETTER S WITH CIRCUMFLEX	Ŝ
U+015D	LATIN SMALL LETTER S WITH CIRCUMFLEX	ŝ
U+015E	LATIN CAPITAL LETTER S WITH CEDILLA	Ș
U+015F	LATIN SMALL LETTER S WITH CEDILLA	ș
U+0160	LATIN CAPITAL LETTER S WITH CARON	Š
U+0161	LATIN SMALL LETTER S WITH CARON	š
U+0162	LATIN CAPITAL LETTER T WITH CEDILLA	Ţ
U+0163	LATIN SMALL LETTER T WITH CEDILLA	ţ
U+0164	LATIN CAPITAL LETTER T WITH CARON	Ť
U+0165	LATIN SMALL LETTER T WITH CARON	ť
U+0166	LATIN CAPITAL LETTER T WITH STROKE	Ƨ
U+0167	LATIN SMALL LETTER T WITH STROKE	Ƨ
U+0168	LATIN CAPITAL LETTER U WITH TILDE	Ũ
U+0169	LATIN SMALL LETTER U WITH TILDE	ũ
U+016A	LATIN CAPITAL LETTER U WITH MACRON	Ū
U+016B	LATIN SMALL LETTER U WITH MACRON	ū
U+016C	LATIN CAPITAL LETTER U WITH BREVE	Ŭ
U+016D	LATIN SMALL LETTER U WITH BREVE	ŭ
U+016E	LATIN CAPITAL LETTER U WITH RING ABOVE	Ů
U+016F	LATIN SMALL LETTER U WITH RING ABOVE	ů
U+0170	LATIN CAPITAL LETTER U WITH DOUBLE ACUTE	Ű
U+0171	LATIN SMALL LETTER U WITH DOUBLE ACUTE	ű
U+0172	LATIN CAPITAL LETTER U WITH OGONEK	Ų
U+0173	LATIN SMALL LETTER U WITH OGONEK	ų
U+0174	LATIN CAPITAL LETTER W WITH CIRCUMFLEX	Ŵ
U+0175	LATIN SMALL LETTER W WITH CIRCUMFLEX	ŵ
U+0176	LATIN CAPITAL LETTER Y WITH CIRCUMFLEX	Ŷ
U+0177	LATIN SMALL LETTER Y WITH CIRCUMFLEX	ŷ
U+0178	LATIN CAPITAL LETTER Y WITH DIAERESIS	ÿ

Codepoints	Name	Glyphe
U+0179	LATIN CAPITAL LETTER Z WITH ACUTE	Ž
U+017A	LATIN SMALL LETTER Z WITH ACUTE	ž
U+017B	LATIN CAPITAL LETTER Z WITH DOT ABOVE	Ž̇
U+017C	LATIN SMALL LETTER Z WITH DOT ABOVE	ž̇
U+017D	LATIN CAPITAL LETTER Z WITH CARON	Ẓ̌
<U+017D,U+0326>	LATIN CAPITAL LETTER Z WITH CARON WITH COMBINING COMMA BELOW	Ẓ̸̌
<U+017D,U+0327>	LATIN CAPITAL LETTER Z WITH CARON WITH COMBINING CEDILLA	Ẓ̧̌
U+017E	LATIN SMALL LETTER Z WITH CARON	ẓ̌
<U+017E,U+0326>	LATIN SMALL LETTER Z WITH CARON WITH COMBINING COMMA BELOW	ẓ̸̌
<U+017E,U+0327>	LATIN SMALL LETTER Z WITH CARON WITH COMBINING CEDILLA	ẓ̧̌
U+0187	LATIN CAPITAL LETTER C WITH HOOK	Ć
U+0188	LATIN SMALL LETTER C WITH HOOK	ć
U+018F	LATIN CAPITAL LETTER SCHWA	Ə
U+0197	LATIN CAPITAL LETTER I WITH STROKE	Ƨ
U+01A0	LATIN CAPITAL LETTER O WITH HORN	Ɔ
U+01A1	LATIN SMALL LETTER O WITH HORN	ɔ
U+01AF	LATIN CAPITAL LETTER U WITH HORN	Ʊ
U+01B0	LATIN SMALL LETTER U WITH HORN	ʊ
U+01B7	LATIN CAPITAL LETTER EZH	Ʒ
U+01CD	LATIN CAPITAL LETTER A WITH CARON	Ā
U+01CE	LATIN SMALL LETTER A WITH CARON	ā
U+01CF	LATIN CAPITAL LETTER I WITH CARON	Ī
U+01D0	LATIN SMALL LETTER I WITH CARON	ī
U+01D1	LATIN CAPITAL LETTER O WITH CARON	Ō
U+01D2	LATIN SMALL LETTER O WITH CARON	ō
U+01D3	LATIN CAPITAL LETTER U WITH CARON	Ū
U+01D4	LATIN SMALL LETTER U WITH CARON	ū
U+01D5	LATIN CAPITAL LETTER U WITH DIAERESIS AND MACRON	Ŭ
U+01D6	LATIN SMALL LETTER U WITH DIAERESIS AND MACRON	ŭ
U+01D7	LATIN CAPITAL LETTER U WITH DIAERESIS AND ACUTE	Ú
U+01D8	LATIN SMALL LETTER U WITH DIAERESIS AND ACUTE	ú
U+01D9	LATIN CAPITAL LETTER U WITH DIAERESIS AND CARON	Ů
U+01DA	LATIN SMALL LETTER U WITH DIAERESIS AND CARON	ů
U+01DB	LATIN CAPITAL LETTER U WITH DIAERESIS AND GRAVE	Û
U+01DC	LATIN SMALL LETTER U WITH DIAERESIS AND GRAVE	û
U+01DE	LATIN CAPITAL LETTER A WITH DIAERESIS AND MACRON	Ä
U+01DF	LATIN SMALL LETTER A WITH DIAERESIS AND MACRON	ä
U+01E2	LATIN CAPITAL LETTER AE WITH MACRON	Æ

Codepoints	Name	Glyphe
U+01E3	LATIN SMALL LETTER AE WITH MACRON	ǣ
U+01E4	LATIN CAPITAL LETTER G WITH STROKE	Ġ
U+01E5	LATIN SMALL LETTER G WITH STROKE	ġ
U+01E6	LATIN CAPITAL LETTER G WITH CARON	Ĝ
U+01E7	LATIN SMALL LETTER G WITH CARON	ĝ
U+01E8	LATIN CAPITAL LETTER K WITH CARON	Ķ
U+01E9	LATIN SMALL LETTER K WITH CARON	ķ
U+01EA	LATIN CAPITAL LETTER O WITH OGONEK	Q̇
U+01EB	LATIN SMALL LETTER O WITH OGONEK	q̇
U+01EC	LATIN CAPITAL LETTER O WITH OGONEK AND MACRON	Q̄̇
U+01ED	LATIN SMALL LETTER O WITH OGONEK AND MACRON	q̄̇
U+01EE	LATIN CAPITAL LETTER EZH WITH CARON	Š
U+01EF	LATIN SMALL LETTER EZH WITH CARON	š
U+01F0	LATIN SMALL LETTER J WITH CARON	Ĵ
U+01F4	LATIN CAPITAL LETTER G WITH ACUTE	Ĝ
U+01F5	LATIN SMALL LETTER G WITH ACUTE	ĝ
U+01F8	LATIN CAPITAL LETTER N WITH GRAVE	Ñ
U+01F9	LATIN SMALL LETTER N WITH GRAVE	ñ
U+01FA	LATIN CAPITAL LETTER A WITH RING ABOVE AND ACUTE	Ą
U+01FB	LATIN SMALL LETTER A WITH RING ABOVE AND ACUTE	ą
U+01FC	LATIN CAPITAL LETTER AE WITH ACUTE	Æ
U+01FD	LATIN SMALL LETTER AE WITH ACUTE	æ
U+01FE	LATIN CAPITAL LETTER O WITH STROKE AND ACUTE	Ø
U+01FF	LATIN SMALL LETTER O WITH STROKE AND ACUTE	ø
U+0212	LATIN CAPITAL LETTER R WITH INVERTED BREVE	Ŕ
U+0213	LATIN SMALL LETTER R WITH INVERTED BREVE	ŕ
U+0218	LATIN CAPITAL LETTER S WITH COMMA BELOW	Ș
U+0219	LATIN SMALL LETTER S WITH COMMA BELOW	ș
U+021A	LATIN CAPITAL LETTER T WITH COMMA BELOW	Ț
U+021B	LATIN SMALL LETTER T WITH COMMA BELOW	ț
U+021E	LATIN CAPITAL LETTER H WITH CARON	Ĥ
U+021F	LATIN SMALL LETTER H WITH CARON	ĥ
U+0227	LATIN SMALL LETTER A WITH DOT ABOVE	ȁ
U+0228	LATIN CAPITAL LETTER E WITH CEDILLA	Ɛ
U+0229	LATIN SMALL LETTER E WITH CEDILLA	ɛ̃
U+022A	LATIN CAPITAL LETTER O WITH DIAERESIS AND MACRON	Ȫ
U+022B	LATIN SMALL LETTER O WITH DIAERESIS AND MACRON	ȫ
U+022C	LATIN CAPITAL LETTER O WITH TILDE AND MACRON	Ȭ
U+022D	LATIN SMALL LETTER O WITH TILDE AND MACRON	ȭ

Codepoints	Name	Glyphe
U+022E	LATIN CAPITAL LETTER O WITH DOT ABOVE	Ŏ
U+022F	LATIN SMALL LETTER O WITH DOT ABOVE	ŏ
U+0230	LATIN CAPITAL LETTER O WITH DOT ABOVE AND MACRON	Ȭ
U+0231	LATIN SMALL LETTER O WITH DOT ABOVE AND MACRON	ȭ
U+0232	LATIN CAPITAL LETTER Y WITH MACRON	Ȳ
U+0233	LATIN SMALL LETTER Y WITH MACRON	ȳ
U+0259	LATIN SMALL LETTER SCHWA	ə
U+0268	LATIN SMALL LETTER I WITH STROKE	ɨ
U+0292	LATIN SMALL LETTER EZH	ɜ
U+1E02	LATIN CAPITAL LETTER B WITH DOT ABOVE	Ḃ
U+1E03	LATIN SMALL LETTER B WITH DOT ABOVE	ḃ
U+1E06	LATIN CAPITAL LETTER B WITH LINE BELOW	Ḅ
U+1E07	LATIN SMALL LETTER B WITH LINE BELOW	ḅ
U+1E0A	LATIN CAPITAL LETTER D WITH DOT ABOVE	Ḍ
U+1E0B	LATIN SMALL LETTER D WITH DOT ABOVE	ḍ
U+1E0C	LATIN CAPITAL LETTER D WITH DOT BELOW	Ḑ
U+1E0D	LATIN SMALL LETTER D WITH DOT BELOW	ḑ
U+1E0E	LATIN CAPITAL LETTER D WITH LINE BELOW	Ḓ
U+1E0F	LATIN SMALL LETTER D WITH LINE BELOW	ḓ
U+1E10	LATIN CAPITAL LETTER D WITH CEDILLA	Ḕ
U+1E11	LATIN SMALL LETTER D WITH CEDILLA	ḕ
U+1E1C	LATIN CAPITAL LETTER E WITH CEDILLA AND BREVE	Ḝ
U+1E1D	LATIN SMALL LETTER E WITH CEDILLA AND BREVE	ḝ
U+1E1E	LATIN CAPITAL LETTER F WITH DOT ABOVE	Ḟ
U+1E1F	LATIN SMALL LETTER F WITH DOT ABOVE	ḟ
U+1E20	LATIN CAPITAL LETTER G WITH MACRON	Ḡ
U+1E21	LATIN SMALL LETTER G WITH MACRON	ḡ
U+1E22	LATIN CAPITAL LETTER H WITH DOT ABOVE	Ḩ
U+1E23	LATIN SMALL LETTER H WITH DOT ABOVE	ḩ
U+1E24	LATIN CAPITAL LETTER H WITH DOT BELOW	Ḭ
U+1E25	LATIN SMALL LETTER H WITH DOT BELOW	ḭ
U+1E26	LATIN CAPITAL LETTER H WITH DIAERESIS	Ḟ
U+1E27	LATIN SMALL LETTER H WITH DIAERESIS	ḟ
U+1E28	LATIN CAPITAL LETTER H WITH CEDILLA	Ḡ
U+1E29	LATIN SMALL LETTER H WITH CEDILLA	ḡ
U+1E2A	LATIN CAPITAL LETTER H WITH BREVE BELOW	Ḣ
U+1E2B	LATIN SMALL LETTER H WITH BREVE BELOW	ḣ
U+1E2F	LATIN SMALL LETTER I WITH DIAERESIS AND ACUTE	í
U+1E30	LATIN CAPITAL LETTER K WITH ACUTE	Ć

Codepoints	Name	Glyphe
U+1E31	LATIN SMALL LETTER K WITH ACUTE	ĳ
U+1E32	LATIN CAPITAL LETTER K WITH DOT BELOW	Ƙ
<U+1E32,U+0304>	LATIN CAPITAL LETTER K WITH DOT BELOW WITH COMBINING MACRON	Ƙ̄
U+1E33	LATIN SMALL LETTER K WITH DOT BELOW	ƙ
<U+1E33,U+0304>	LATIN SMALL LETTER K WITH DOT BELOW WITH COMBINING MACRON	ƙ̄
U+1E34	LATIN CAPITAL LETTER K WITH LINE BELOW	ƚ
U+1E35	LATIN SMALL LETTER K WITH LINE BELOW	ƚ
U+1E36	LATIN CAPITAL LETTER L WITH DOT BELOW	ƚ
U+1E37	LATIN SMALL LETTER L WITH DOT BELOW	ƚ
U+1E3A	LATIN CAPITAL LETTER L WITH LINE BELOW	ƚ
U+1E3B	LATIN SMALL LETTER L WITH LINE BELOW	ƚ
U+1E40	LATIN CAPITAL LETTER M WITH DOT ABOVE	Ṁ
U+1E41	LATIN SMALL LETTER M WITH DOT ABOVE	ṁ
U+1E42	LATIN CAPITAL LETTER M WITH DOT BELOW	Ṃ
U+1E43	LATIN SMALL LETTER M WITH DOT BELOW	ṃ
U+1E44	LATIN CAPITAL LETTER N WITH DOT ABOVE	Ṇ
U+1E45	LATIN SMALL LETTER N WITH DOT ABOVE	ṇ
U+1E46	LATIN CAPITAL LETTER N WITH DOT BELOW	Ṉ
U+1E47	LATIN SMALL LETTER N WITH DOT BELOW	ṉ
U+1E48	LATIN CAPITAL LETTER N WITH LINE BELOW	Ṋ
U+1E49	LATIN SMALL LETTER N WITH LINE BELOW	ṋ
U+1E52	LATIN CAPITAL LETTER O WITH MACRON AND ACUTE	Ȫ
U+1E53	LATIN SMALL LETTER O WITH MACRON AND ACUTE	ȫ
U+1E54	LATIN CAPITAL LETTER P WITH ACUTE	Ṗ
U+1E55	LATIN SMALL LETTER P WITH ACUTE	ṗ
U+1E56	LATIN CAPITAL LETTER P WITH DOT ABOVE	Ṙ
U+1E57	LATIN SMALL LETTER P WITH DOT ABOVE	ṙ
U+1E58	LATIN CAPITAL LETTER R WITH DOT ABOVE	Ṛ
U+1E59	LATIN SMALL LETTER R WITH DOT ABOVE	ṓ
U+1E5A	LATIN CAPITAL LETTER R WITH DOT BELOW	Ṕ
U+1E5B	LATIN SMALL LETTER R WITH DOT BELOW	ṑ
U+1E5E	LATIN CAPITAL LETTER R WITH LINE BELOW	Ṗ
U+1E5F	LATIN SMALL LETTER R WITH LINE BELOW	ṗ
U+1E60	LATIN CAPITAL LETTER S WITH DOT ABOVE	Ṣ
U+1E61	LATIN SMALL LETTER S WITH DOT ABOVE	ṣ
U+1E62	LATIN CAPITAL LETTER S WITH DOT BELOW	Ṥ
<U+1E62,U+0304>	LATIN CAPITAL LETTER S WITH DOT BELOW WITH COMBINING MACRON	Ṥ̄

Codepoints	Name	Glyphe
U+1E63	LATIN SMALL LETTER S WITH DOT BELOW	ŝ
<U+1E63,U+0304>	LATIN SMALL LETTER S WITH DOT BELOW WITH COMBINING MACRON	ŝ̄
U+1E6A	LATIN CAPITAL LETTER T WITH DOT ABOVE	Ṫ
U+1E6B	LATIN SMALL LETTER T WITH DOT ABOVE	ṭ
U+1E6C	LATIN CAPITAL LETTER T WITH DOT BELOW	Ṭ
<U+1E6C,U+0304>	LATIN CAPITAL LETTER T WITH DOT BELOW WITH COMBINING MACRON	Ṭ̄
U+1E6D	LATIN SMALL LETTER T WITH DOT BELOW	ṭ
<U+1E6D,U+0304>	LATIN SMALL LETTER T WITH DOT BELOW WITH COMBINING MACRON	ṭ̄
U+1E6E	LATIN CAPITAL LETTER T WITH LINE BELOW	Ṛ
U+1E6F	LATIN SMALL LETTER T WITH LINE BELOW	ṛ
U+1E80	LATIN CAPITAL LETTER W WITH GRAVE	Ẁ
U+1E81	LATIN SMALL LETTER W WITH GRAVE	ẁ
U+1E82	LATIN CAPITAL LETTER W WITH ACUTE	Ẃ
U+1E83	LATIN SMALL LETTER W WITH ACUTE	ẃ
U+1E84	LATIN CAPITAL LETTER W WITH DIAERESIS	Ẅ
U+1E85	LATIN SMALL LETTER W WITH DIAERESIS	ẅ
U+1E86	LATIN CAPITAL LETTER W WITH DOT ABOVE	Ẇ
U+1E87	LATIN SMALL LETTER W WITH DOT ABOVE	ẇ
U+1E8C	LATIN CAPITAL LETTER X WITH DIAERESIS	Ẋ
U+1E8D	LATIN SMALL LETTER X WITH DIAERESIS	ẋ
U+1E8E	LATIN CAPITAL LETTER Y WITH DOT ABOVE	Ỳ
U+1E8F	LATIN SMALL LETTER Y WITH DOT ABOVE	ỳ
U+1E90	LATIN CAPITAL LETTER Z WITH CIRCUMFLEX	Ẑ
U+1E91	LATIN SMALL LETTER Z WITH CIRCUMFLEX	ẑ
U+1E92	LATIN CAPITAL LETTER Z WITH DOT BELOW	Ẓ
U+1E93	LATIN SMALL LETTER Z WITH DOT BELOW	ẓ
U+1E94	LATIN CAPITAL LETTER Z WITH LINE BELOW	Ẕ
U+1E95	LATIN SMALL LETTER Z WITH LINE BELOW	ẕ
U+1E96	LATIN SMALL LETTER H WITH LINE BELOW	ḥ
U+1E97	LATIN SMALL LETTER T WITH DIAERESIS	ṫ
U+1E9E	LATIN CAPITAL LETTER SHARP S	ß
U+1EA0	LATIN CAPITAL LETTER A WITH DOT BELOW	Ạ
<U+1EA0,U+0308>	LATIN CAPITAL LETTER A WITH DOT BELOW WITH COMBINING DIAERESIS	Ạ̈
U+1EA1	LATIN SMALL LETTER A WITH DOT BELOW	ạ
<U+1EA1,U+0308>	LATIN SMALL LETTER A WITH DOT BELOW WITH COMBINING DIAERESIS	ạ̈
U+1EA2	LATIN CAPITAL LETTER A WITH HOOK ABOVE	Ả

Codepoints	Name	Glyphe
U+1EA3	LATIN SMALL LETTER A WITH HOOK ABOVE	ǎ
U+1EA4	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND ACUTE	Ấ
U+1EA5	LATIN SMALL LETTER A WITH CIRCUMFLEX AND ACUTE	ấ
U+1EA6	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND GRAVE	Ầ
U+1EA7	LATIN SMALL LETTER A WITH CIRCUMFLEX AND GRAVE	ầ
U+1EA8	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE	Ấ̃
U+1EA9	LATIN SMALL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE	ấ̃
U+1EAA	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND TILDE	Ẫ
U+1EAB	LATIN SMALL LETTER A WITH CIRCUMFLEX AND TILDE	ẫ
U+1EAC	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND DOT BELOW	Ậ
U+1EAD	LATIN SMALL LETTER A WITH CIRCUMFLEX AND DOT BELOW	ậ
U+1EAE	LATIN CAPITAL LETTER A WITH BREVE AND ACUTE	Ắ
U+1EAF	LATIN SMALL LETTER A WITH BREVE AND ACUTE	ắ
U+1EB0	LATIN CAPITAL LETTER A WITH BREVE AND GRAVE	Ằ
U+1EB1	LATIN SMALL LETTER A WITH BREVE AND GRAVE	ằ
U+1EB2	LATIN CAPITAL LETTER A WITH BREVE AND HOOK ABOVE	Ắ̃
U+1EB3	LATIN SMALL LETTER A WITH BREVE AND HOOK ABOVE	ắ̃
U+1EB4	LATIN CAPITAL LETTER A WITH BREVE AND TILDE	Ẵ
U+1EB5	LATIN SMALL LETTER A WITH BREVE AND TILDE	ẵ
U+1EB6	LATIN CAPITAL LETTER A WITH BREVE AND DOT BELOW	Ặ
U+1EB7	LATIN SMALL LETTER A WITH BREVE AND DOT BELOW	ặ
U+1EB8	LATIN CAPITAL LETTER E WITH DOT BELOW	Ț
U+1EB9	LATIN SMALL LETTER E WITH DOT BELOW	Ț̣
U+1EBA	LATIN CAPITAL LETTER E WITH HOOK ABOVE	Ế
U+1EBB	LATIN SMALL LETTER E WITH HOOK ABOVE	ế
U+1EBC	LATIN CAPITAL LETTER E WITH TILDE	Ễ
U+1EBD	LATIN SMALL LETTER E WITH TILDE	ễ
U+1EBE	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND ACUTE	Ế̂
U+1EBF	LATIN SMALL LETTER E WITH CIRCUMFLEX AND ACUTE	ế̂
U+1EC0	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND GRAVE	Ề̂
U+1EC1	LATIN SMALL LETTER E WITH CIRCUMFLEX AND GRAVE	ề̂
U+1EC2	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE	Ế̂̃
U+1EC3	LATIN SMALL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE	ế̂̃
U+1EC4	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND TILDE	Ê̂̃
U+1EC5	LATIN SMALL LETTER E WITH CIRCUMFLEX AND TILDE	ê̂̃
U+1EC6	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND DOT BELOW	Ệ̂
U+1EC7	LATIN SMALL LETTER E WITH CIRCUMFLEX AND DOT BELOW	ệ̂
U+1EC8	LATIN CAPITAL LETTER I WITH HOOK ABOVE	İ́
U+1EC9	LATIN SMALL LETTER I WITH HOOK ABOVE	ı́

Codepoints	Name	Glyphe
U+1ECA	LATIN CAPITAL LETTER I WITH DOT BELOW	İ
U+1ECB	LATIN SMALL LETTER I WITH DOT BELOW	ı
U+1ECC	LATIN CAPITAL LETTER O WITH DOT BELOW	Ȯ
<U+1ECC,U+0308>	LATIN CAPITAL LETTER O WITH DOT BELOW WITH COMBINING DIAERESIS	Ö
U+1ECD	LATIN SMALL LETTER O WITH DOT BELOW	ȯ
<U+1ECD,U+0308>	LATIN SMALL LETTER O WITH DOT BELOW WITH COMBINING DIAERESIS	ö
U+1ECE	LATIN CAPITAL LETTER O WITH HOOK ABOVE	Ȫ
U+1ECF	LATIN SMALL LETTER O WITH HOOK ABOVE	ȫ
U+1ED0	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND ACUTE	Ô
U+1ED1	LATIN SMALL LETTER O WITH CIRCUMFLEX AND ACUTE	ô
U+1ED2	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND GRAVE	Ȫ
U+1ED3	LATIN SMALL LETTER O WITH CIRCUMFLEX AND GRAVE	ȫ
U+1ED4	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE	Ȫ
U+1ED5	LATIN SMALL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE	ȫ
U+1ED6	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND TILDE	Ȫ
U+1ED7	LATIN SMALL LETTER O WITH CIRCUMFLEX AND TILDE	ȫ
U+1ED8	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND DOT BELOW	Ȫ
U+1ED9	LATIN SMALL LETTER O WITH CIRCUMFLEX AND DOT BELOW	ȫ
U+1EDA	LATIN CAPITAL LETTER O WITH HORN AND ACUTE	Ȫ
U+1EDB	LATIN SMALL LETTER O WITH HORN AND ACUTE	ȫ
U+1EDC	LATIN CAPITAL LETTER O WITH HORN AND GRAVE	Ȫ
U+1EDD	LATIN SMALL LETTER O WITH HORN AND GRAVE	ȫ
U+1EDE	LATIN CAPITAL LETTER O WITH HORN AND HOOK ABOVE	Ȫ
U+1EDF	LATIN SMALL LETTER O WITH HORN AND HOOK ABOVE	ȫ
U+1EE0	LATIN CAPITAL LETTER O WITH HORN AND TILDE	Ȫ
U+1EE1	LATIN SMALL LETTER O WITH HORN AND TILDE	ȫ
U+1EE2	LATIN CAPITAL LETTER O WITH HORN AND DOT BELOW	Ȫ
U+1EE3	LATIN SMALL LETTER O WITH HORN AND DOT BELOW	ȫ
U+1EE4	LATIN CAPITAL LETTER U WITH DOT BELOW	Ȫ
<U+1EE4,U+0304>	LATIN CAPITAL LETTER U WITH DOT BELOW WITH COMBINING MACRON	Ȫ
<U+1EE4,U+0308>	LATIN CAPITAL LETTER U WITH DOT BELOW WITH COMBINING DIAERESIS	Ȫ
U+1EE5	LATIN SMALL LETTER U WITH DOT BELOW	ȫ
<U+1EE5,U+0304>	LATIN SMALL LETTER U WITH DOT BELOW WITH COMBINING MACRON	ȫ
<U+1EE5,U+0308>	LATIN SMALL LETTER U WITH DOT BELOW WITH COMBINING DIAERESIS	ȫ
U+1EE6	LATIN CAPITAL LETTER U WITH HOOK ABOVE	Ȫ

Codepoints	Name	Glyphe
U+1EE7	LATIN SMALL LETTER U WITH HOOK ABOVE	ŭ
U+1EE8	LATIN CAPITAL LETTER U WITH HORN AND ACUTE	Ū́
U+1EE9	LATIN SMALL LETTER U WITH HORN AND ACUTE	ū́
U+1EEA	LATIN CAPITAL LETTER U WITH HORN AND GRAVE	Ū̀
U+1EEB	LATIN SMALL LETTER U WITH HORN AND GRAVE	ū̀
U+1EEC	LATIN CAPITAL LETTER U WITH HORN AND HOOK ABOVE	Ū̂
U+1EED	LATIN SMALL LETTER U WITH HORN AND HOOK ABOVE	ū̂
U+1EEE	LATIN CAPITAL LETTER U WITH HORN AND TILDE	Ū̃
U+1EEF	LATIN SMALL LETTER U WITH HORN AND TILDE	ū̃
U+1EF0	LATIN CAPITAL LETTER U WITH HORN AND DOT BELOW	Ụ̄
U+1EF1	LATIN SMALL LETTER U WITH HORN AND DOT BELOW	ụ̄
U+1EF2	LATIN CAPITAL LETTER Y WITH GRAVE	Ỳ
U+1EF3	LATIN SMALL LETTER Y WITH GRAVE	ỳ
U+1EF4	LATIN CAPITAL LETTER Y WITH DOT BELOW	Ỵ
U+1EF5	LATIN SMALL LETTER Y WITH DOT BELOW	ỵ
U+1EF6	LATIN CAPITAL LETTER Y WITH HOOK ABOVE	Ỳ̂
U+1EF7	LATIN SMALL LETTER Y WITH HOOK ABOVE	ỳ̂
U+1EF8	LATIN CAPITAL LETTER Y WITH TILDE	Ỳ̃
U+1EF9	LATIN SMALL LETTER Y WITH TILDE	ỳ̃

4.3. Kombinierende diakritische Zeichen

Die in Tabelle 4 aufgeführten „kombinierenden diakritischen Zeichen“ werden benötigt, um die in der Tabelle 3 genannten Zeichensequenzen zu bilden.

Tabelle 4. Normative kombinierende diakritische Zeichen

Codepoints	Name	Glyphe
U+0300	COMBINING GRAVE ACCENT	`
U+0301	COMBINING ACUTE ACCENT	´
U+0302	COMBINING CIRCUMFLEX ACCENT	^
U+0304	COMBINING MACRON	¯
U+0306	COMBINING BREVE	˘
U+0307	COMBINING DOT ABOVE	˙
U+0308	COMBINING DIAERESIS	¨
U+030B	COMBINING DOUBLE ACUTE ACCENT	˝
U+030C	COMBINING CARON	ˇ
U+0310	COMBINING CANDRABINDU	◌̣
U+0315	COMBINING COMMA ABOVE RIGHT	◌̵
U+031B	COMBINING HORN	◌̶
U+0323	COMBINING DOT BELOW	◌̣

Codepoints	Name	Glyphe
U+0325	COMBINING RING BELOW	◌◌
U+0326	COMBINING COMMA BELOW	◌,
U+0327	COMBINING CEDILLA	◌¸
U+0328	COMBINING OGONEK	◌˛
U+0331	COMBINING MACRON BELOW	◌˘
U+035F	COMBINING DOUBLE MACRON BELOW	◌˝

Kombinierende diakritische Zeichen bedürfen bei der in Schnittstellenvereinbarungen erforderlichen Einschränkung eines Zeichensatzes besonderer Beachtung. Denn in beliebiger Kombination mit Grundzeichen könnten sie genutzt werden um Zeichen zu bilden, die eigentlich ausgeschlossen sein sollten.

Daher wurden die Datentypen des Abschnitts B so definiert, dass Kombinierende diakritische Zeichen nur in den in Tabelle 3 genannten Sequenzen zulässig sind. Das setzt voraus, dass die Daten in der Unicode Normalform NFC ausgetauscht werden, siehe Abschnitt Unicode Normalisierung NFC.

4.4. Nicht-Buchstaben N1

Die in Tabelle 5 genannten Schriftzeichen sind für die Darstellung von allen Namen im weiteren Sinne, insbesondere für Namen natürlicher Personen, zwingend erforderlich.

Tabelle 5. Normative Nicht-Buchstaben, erforderlich für alle Namen

Codepoints	Name	Glyphe
U+0020	SPACE	
U+0027	APOSTROPHE	'
U+002C	COMMA	,
U+002D	HYPHEN-MINUS	-
U+002E	FULL STOP	.
U+0060	GRAVE ACCENT	`
U+007E	TILDE	~
U+00A8	DIAERESIS	¨
U+00B4	ACUTE ACCENT	´
U+00B7	MIDDLE DOT	·
U+02B9	MODIFIER LETTER PRIME	′
U+02BA	MODIFIER LETTER DOUBLE PRIME	″
U+02BE	MODIFIER LETTER RIGHT HALF RING	ʹ
U+02BF	MODIFIER LETTER LEFT HALF RING	ʻ
U+02C8	MODIFIER LETTER VERTICAL LINE	⋈
U+02CC	MODIFIER LETTER LOW VERTICAL LINE	⋚
U+2019	RIGHT SINGLE QUOTATION MARK	’
U+2021	DOUBLE DAGGER	‡

4.5. Nicht-Buchstaben N2

Die in Tabelle 6 genannten Schriftzeichen sind für die Darstellung von vielen Namen im weiteren Sinne, z. B. für Ortsnamen, für die Namen von Straßen inkl. Hausnummer, für die Namen juristischer Personen und für Produktnamen, zwingend erforderlich. Dies gilt jedoch nicht für alle Namen, insbesondere nicht für Namen natürlicher Personen.

Tabelle 6. Sonstige normative Nicht-Buchstaben, erforderlich für nicht alle Namen

Codepoints	Name	Glyphe
U+0021	EXCLAMATION MARK	!
U+0022	QUOTATION MARK	"
U+0023	NUMBER SIGN	#
U+0024	DOLLAR SIGN	\$
U+0025	PERCENT SIGN	%
U+0026	AMPERSAND	&
U+0028	LEFT PARENTHESIS	(
U+0029	RIGHT PARENTHESIS)
U+002A	ASTERISK	*
U+002B	PLUS SIGN	+
U+002F	SOLIDUS	/
U+0030	DIGIT ZERO	0
U+0031	DIGIT ONE	1
U+0032	DIGIT TWO	2
U+0033	DIGIT THREE	3
U+0034	DIGIT FOUR	4
U+0035	DIGIT FIVE	5
U+0036	DIGIT SIX	6
U+0037	DIGIT SEVEN	7
U+0038	DIGIT EIGHT	8
U+0039	DIGIT NINE	9
U+003A	COLON	:
U+003B	SEMICOLON	;
U+003C	LESS-THAN SIGN	<
U+003D	EQUALS SIGN	=
U+003E	GREATER-THAN SIGN	>
U+003F	QUESTION MARK	?
U+0040	COMMERCIAL AT	@
U+005B	LEFT SQUARE BRACKET	[
U+005C	REVERSE SOLIDUS	\
U+005D	RIGHT SQUARE BRACKET]
U+005E	CIRCUMFLEX ACCENT	^
U+005F	LOW LINE	—

Codepoints	Name	Glyphe
U+007B	LEFT CURLY BRACKET	{
U+007C	VERTICAL LINE	
U+007D	RIGHT CURLY BRACKET	}
U+00A1	INVERTED EXCLAMATION MARK	¡
U+00A2	CENT SIGN	¢
U+00A3	POUND SIGN	£
U+00A5	YEN SIGN	¥
U+00A7	SECTION SIGN	§
U+00A9	COPYRIGHT SIGN	©
U+00AA	FEMININE ORDINAL INDICATOR	ª
U+00AB	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK	«
U+00AC	NOT SIGN	¬
U+00AE	REGISTERED SIGN	®
U+00AF	MACRON	-
U+00B0	DEGREE SIGN	°
U+00B1	PLUS-MINUS SIGN	±
U+00B2	SUPERSCRIFT TWO	²
U+00B3	SUPERSCRIFT THREE	³
U+00B5	MICRO SIGN	μ
U+00B6	PILCROW SIGN	¶
U+00B9	SUPERSCRIFT ONE	¹
U+00BA	MASCULINE ORDINAL INDICATOR	º
U+00BB	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK	»
U+00BF	INVERTED QUESTION MARK	¿
U+00D7	MULTIPLICATION SIGN	×
U+00F7	DIVISION SIGN	÷
U+20AC	EURO SIGN	€

4.6. Nicht-Buchstaben N3

Die in Tabelle 7 genannten Schriftzeichen sind wegen der Abwärtskompatibilität zum Standard [String Latin 1.1] zwingend erforderlich. Für die Darstellung von Namen natürlicher Personen sowie sonstiger Namen, wie z. B. Ortsnamen und Straßennamen, sind diese Zeichen ungeeignet und sollten deshalb bei allen für diese Namen bestimmten Datenfeldern nicht zum Einsatz kommen. Diese Zeichen sind jedoch relevant für Namen juristischer Personen und für Produktnamen.

Tabelle 7. Normative Nicht-Buchstaben, erforderlich für Namen juristischer Personen

Codepoints	Name	Glyphe
U+00A4	CURRENCY SIGN	¤
U+00A6	BROKEN BAR	
U+00B8	CEDILLA	¸

Codepoints	Name	Glyphe
U+00BC	VULGAR FRACTION ONE QUARTER	¼
U+00BD	VULGAR FRACTION ONE HALF	½
U+00BE	VULGAR FRACTION THREE QUARTERS	¾

4.7. Nicht-Buchstaben N4

Die in Tabelle 8 genannten Schriftzeichen sind für die Darstellung von Namen (im weiteren Sinne) in der Regel ungeeignet und sollten deshalb bei für Namen bestimmten Datenfeldern nicht zum Einsatz kommen. Nichtsdestotrotz müssen IT-Systeme im Stande sein, bei Bedarf auch diese Zeichen in vollem Umfang zu verarbeiten. Es kann für spezielle Namen notwendig sein, alle Zeichen in einer Zeile zu halten, weil ein Zeilenumbruch den Sinn entstellen könnte oder Namensbestandteile aus dem Zusammenhang reißen könnte. Deshalb müssen zur DIN SPEC konforme Systeme das „NO-BREAK SPACE“ unterstützen. Die anderen Zeichen befinden sich im normativen Teil, um die Abwärtskompatibilität zum Standard „Lateinische Zeichen in Unicode Version 1.1“ (auch „String.Latin 1.1“ genannt) sicherzustellen.

Tabelle 8. Normative Nicht-Buchstaben, für Namen ungeeignet

Codepoints	Name	Glyphe
U+0009	CHARACTER TABULATION	␣
U+000A	LINE FEED (LF)	␣
U+000D	CARRIAGE RETURN (CR)	␣
U+00A0	NO-BREAK SPACE	

Kapitel 5. Normative Abbildung lateinischer Buchstaben auf Grundbuchstaben (Suchform)

Für unterschiedliche Anwendungsfälle kann es erforderlich sein, die in Namen enthaltenen Buchstaben auf die Grundbuchstaben A bis Z abzubilden. Insbesondere stehen in der maschinenlesbaren Zone (MRZ) amtlicher Reisedokumente, für die der Standard [ICAO 9303-3] einschlägig ist, nur diese Grundbuchstaben für die Repräsentation von Namen zur Verfügung.

Die Tabelle 9 legt für alle in dieser DIN SPEC normativ vorgegebenen Buchstaben deren Abbildung auf die Grundbuchstaben A bis Z fest. Sie basiert auf der Empfehlung zur Transliteration in Abschnitt 6 ("TRANSLITERATIONS RECOMMENDED FOR USE BY STATES") des Dokuments [ICAO 9303-3]. Die darin enthaltenen Transliterationsempfehlungen werden übernommen. Zudem wird eine entsprechende Transliteration auch für diejenigen lateinischen Buchstaben festgelegt, zu denen die ICAO Empfehlung keine Aussage trifft.

Ein weiterer wesentlicher Anwendungsfall ist die Bildung von normalisierten Suchformen, um Namen auch dann in Datenbeständen zu finden, wenn diese noch nicht alle Zeichen der DIN SPEC unterstützen bzw. wenn es Bestandsdaten gibt, bei denen Buchstaben mit diakritischen Zeichen aufgrund technischer Restriktionen von IT-Verfahren ersetzt worden sind. Wenn Datensätze und Suchanfragen unter Anwendung der Suchform normalisiert werden, ist es möglich, Namen trotz unterschiedlicher Schreibweisen zu identifizieren. Dieser Sachverhalt ist im Dokument [AKI Identifikation] näher ausgeführt.

Tabelle 9. Abbildung lateinischer Buchstaben auf Grundbuchstaben analog ICAO

Quelle			Ziel	
U+0041	LATIN CAPITAL LETTER A	A	U+0041	A
<U+0041,U+030B>	LATIN CAPITAL LETTER A WITH COMBINING DOUBLE ACUTE ACCENT	Ǻ	U+0041	A
U+0042	LATIN CAPITAL LETTER B	B	U+0042	B
U+0043	LATIN CAPITAL LETTER C	C	U+0043	C
<U+0043,U+0300>	LATIN CAPITAL LETTER C WITH COMBINING GRAVE ACCENT	Ĉ	U+0043	C
<U+0043,U+0304>	LATIN CAPITAL LETTER C WITH COMBINING MACRON	Ċ	U+0043	C
<U+0043,U+0306>	LATIN CAPITAL LETTER C WITH COMBINING BREVE	Ċ	U+0043	C
<U+0043,U+0308>	LATIN CAPITAL LETTER C WITH COMBINING DIAERESIS	Č	U+0043	C
<U+0043,U+0315>	LATIN CAPITAL LETTER C WITH COMBINING COMMA ABOVE RIGHT	Ć	U+0043	C
<U+0043,U+0323>	LATIN CAPITAL LETTER C WITH COMBINING DOT BELOW	Ċ	U+0043	C
<U+0043,U+0326>	LATIN CAPITAL LETTER C WITH COMBINING COMMA BELOW	Ċ	U+0043	C
<U+0043,U+0328,U+0306>	LATIN CAPITAL LETTER C WITH COMBINING OGONEK AND COMBINING BREVE	Ċ	U+0043	C

Quelle			Ziel	
U+0044	LATIN CAPITAL LETTER D	D	U+0044	D
<U+0044,U+0302>	LATIN CAPITAL LETTER D WITH COMBINING CIRCUMFLEX ACCENT	Ď	U+0044	D
U+0045	LATIN CAPITAL LETTER E	E	U+0045	E
U+0046	LATIN CAPITAL LETTER F	F	U+0046	F
<U+0046,U+0300>	LATIN CAPITAL LETTER F WITH COMBINING GRAVE ACCENT	Ě	U+0046	F
<U+0046,U+0304>	LATIN CAPITAL LETTER F WITH COMBINING MACRON	Ĝ	U+0046	F
U+0047	LATIN CAPITAL LETTER G	G	U+0047	G
<U+0047,U+0300>	LATIN CAPITAL LETTER G WITH COMBINING GRAVE ACCENT	Ĝ	U+0047	G
U+0048	LATIN CAPITAL LETTER H	H	U+0048	H
<U+0048,U+0304>	LATIN CAPITAL LETTER H WITH COMBINING MACRON	Ĥ	U+0048	H
<U+0048,U+0326>	LATIN CAPITAL LETTER H WITH COMBINING COMMA BELOW	Ħ	U+0048	H
<U+0048,U+0331>	LATIN CAPITAL LETTER H WITH COMBINING MACRON BELOW	Ħ	U+0048	H
U+0049	LATIN CAPITAL LETTER I	I	U+0049	I
U+004A	LATIN CAPITAL LETTER J	J	U+004A	J
<U+004A,U+0301>	LATIN CAPITAL LETTER J WITH COMBINING ACUTE ACCENT	Ĵ	U+004A	J
<U+004A,U+030C>	LATIN CAPITAL LETTER J WITH COMBINING CARON	Ĵ	U+004A	J
U+004B	LATIN CAPITAL LETTER K	K	U+004B	K
<U+004B,U+0300>	LATIN CAPITAL LETTER K WITH COMBINING GRAVE ACCENT	Ķ	U+004B	K
<U+004B,U+0302>	LATIN CAPITAL LETTER K WITH COMBINING CIRCUMFLEX ACCENT	Ķ	U+004B	K
<U+004B,U+0304>	LATIN CAPITAL LETTER K WITH COMBINING MACRON	Ķ	U+004B	K
<U+004B,U+0307>	LATIN CAPITAL LETTER K WITH COMBINING DOT ABOVE	Ķ	U+004B	K
<U+004B,U+0315>	LATIN CAPITAL LETTER K WITH COMBINING COMMA ABOVE RIGHT	Ķ	U+004B	K
<U+004B,U+031B>	LATIN CAPITAL LETTER K WITH COMBINING HORN	Ķ	U+004B	K
<U+004B,U+0326>	LATIN CAPITAL LETTER K WITH COMBINING COMMA BELOW	Ķ	U+004B	K
<U+004B,U+035F,U+0048>	LATIN CAPITAL LETTER K WITH COMBINING DOUBLE MACRON BELOW AND LATIN CAPITAL LETTER H	ĶH	<U+004B,U+0048>	KH

Quelle			Ziel	
<U+004B,U+035F,U+0068>	LATIN CAPITAL LETTER K WITH COMBINING DOUBLE MACRON BELOW AND LATIN SMALL LETTER H	Kh	<U+004B,U+0048>	KH
U+004C	LATIN CAPITAL LETTER L	L	U+004C	L
<U+004C,U+0302>	LATIN CAPITAL LETTER L WITH COMBINING CIRCUMFLEX ACCENT	Ĺ	U+004C	L
<U+004C,U+0325>	LATIN CAPITAL LETTER L WITH COMBINING RING BELOW	Ł	U+004C	L
<U+004C,U+0325,U+0304>	LATIN CAPITAL LETTER L WITH COMBINING RING BELOW AND COMBINING MACRON	Ł̄	U+004C	L
<U+004C,U+0326>	LATIN CAPITAL LETTER L WITH COMBINING COMMA BELOW	Ł̸	U+004C	L
U+004D	LATIN CAPITAL LETTER M	M	U+004D	M
<U+004D,U+0300>	LATIN CAPITAL LETTER M WITH COMBINING GRAVE ACCENT	Ṁ	U+004D	M
<U+004D,U+0302>	LATIN CAPITAL LETTER M WITH COMBINING CIRCUMFLEX ACCENT	Ṃ	U+004D	M
<U+004D,U+0306>	LATIN CAPITAL LETTER M WITH COMBINING BREVE	Ṅ	U+004D	M
<U+004D,U+0310>	LATIN CAPITAL LETTER M WITH COMBINING CANDRABINDU	Ṇ	U+004D	M
U+004E	LATIN CAPITAL LETTER N	N	U+004E	N
<U+004E,U+0302>	LATIN CAPITAL LETTER N WITH COMBINING CIRCUMFLEX ACCENT	Ñ	U+004E	N
<U+004E,U+0304>	LATIN CAPITAL LETTER N WITH COMBINING MACRON	Ñ̄	U+004E	N
<U+004E,U+0306>	LATIN CAPITAL LETTER N WITH COMBINING BREVE	Ñ̆	U+004E	N
<U+004E,U+0326>	LATIN CAPITAL LETTER N WITH COMBINING COMMA BELOW	Ñ̸	U+004E	N
U+004F	LATIN CAPITAL LETTER O	O	U+004F	O
U+0050	LATIN CAPITAL LETTER P	P	U+0050	P
<U+0050,U+0300>	LATIN CAPITAL LETTER P WITH COMBINING GRAVE ACCENT	Ṗ	U+0050	P
<U+0050,U+0304>	LATIN CAPITAL LETTER P WITH COMBINING MACRON	Ṗ̄	U+0050	P
<U+0050,U+0315>	LATIN CAPITAL LETTER P WITH COMBINING COMMA ABOVE RIGHT	Ṗ̸	U+0050	P
<U+0050,U+0323>	LATIN CAPITAL LETTER P WITH COMBINING DOT BELOW	Ṗ̣	U+0050	P
U+0051	LATIN CAPITAL LETTER Q	Q	U+0051	Q
U+0052	LATIN CAPITAL LETTER R	R	U+0052	R
<U+0052,U+0306>	LATIN CAPITAL LETTER R WITH COMBINING BREVE	Ṛ	U+0052	R

Quelle			Ziel	
<U+0052,U+0325>	LATIN CAPITAL LETTER R WITH COMBINING RING BELOW	Ṛ	U+0052	R
<U+0052,U+0325,U+0304>	LATIN CAPITAL LETTER R WITH COMBINING RING BELOW AND COMBINING MACRON	Ṝ	U+0052	R
U+0053	LATIN CAPITAL LETTER S	S	U+0053	S
<U+0053,U+0300>	LATIN CAPITAL LETTER S WITH COMBINING GRAVE ACCENT	Ŝ	U+0053	S
<U+0053,U+0304>	LATIN CAPITAL LETTER S WITH COMBINING MACRON	Ŝ̄	U+0053	S
<U+0053,U+031B,U+0304>	LATIN CAPITAL LETTER S WITH COMBINING HORN AND COMBINING MACRON	Ŝ̄̈	U+0053	S
<U+0053,U+0331>	LATIN CAPITAL LETTER S WITH COMBINING MACRON BELOW	Ṣ̣	U+0053	S
U+0054	LATIN CAPITAL LETTER T	T	U+0054	T
<U+0054,U+0300>	LATIN CAPITAL LETTER T WITH COMBINING GRAVE ACCENT	Ť	U+0054	T
<U+0054,U+0304>	LATIN CAPITAL LETTER T WITH COMBINING MACRON	Ť̄	U+0054	T
<U+0054,U+0308>	LATIN CAPITAL LETTER T WITH COMBINING DIAERESIS	Ŧ	U+0054	T
<U+0054,U+0315>	LATIN CAPITAL LETTER T WITH COMBINING COMMA ABOVE RIGHT	Ṭ	U+0054	T
<U+0054,U+031B>	LATIN CAPITAL LETTER T WITH COMBINING HORN	Ṯ	U+0054	T
U+0055	LATIN CAPITAL LETTER U	U	U+0055	U
<U+0055,U+0307>	LATIN CAPITAL LETTER U WITH COMBINING DOT ABOVE	Û	U+0055	U
U+0056	LATIN CAPITAL LETTER V	V	U+0056	V
U+0057	LATIN CAPITAL LETTER W	W	U+0057	W
U+0058	LATIN CAPITAL LETTER X	X	U+0058	X
U+0059	LATIN CAPITAL LETTER Y	Y	U+0059	Y
U+005A	LATIN CAPITAL LETTER Z	Z	U+005A	Z
<U+005A,U+0300>	LATIN CAPITAL LETTER Z WITH COMBINING GRAVE ACCENT	Ẑ	U+005A	Z
<U+005A,U+0304>	LATIN CAPITAL LETTER Z WITH COMBINING MACRON	Ẓ̄	U+005A	Z
<U+005A,U+0306>	LATIN CAPITAL LETTER Z WITH COMBINING BREVE	Ẕ̆	U+005A	Z
<U+005A,U+0308>	LATIN CAPITAL LETTER Z WITH COMBINING DIAERESIS	Ẕ̈	U+005A	Z
<U+005A,U+0327>	LATIN CAPITAL LETTER Z WITH COMBINING CEDILLA	Ẑ̃	U+005A	Z
U+0061	LATIN SMALL LETTER A	a	U+0041	A

Quelle			Ziel	
<U+0061,U+030B>	LATIN SMALL LETTER A WITH COMBINING DOUBLE ACUTE ACCENT	ǎ	U+0041	A
U+0062	LATIN SMALL LETTER B	b	U+0042	B
U+0063	LATIN SMALL LETTER C	c	U+0043	C
<U+0063,U+0300>	LATIN SMALL LETTER C WITH COMBINING GRAVE ACCENT	ç	U+0043	C
<U+0063,U+0304>	LATIN SMALL LETTER C WITH COMBINING MACRON	ċ	U+0043	C
<U+0063,U+0306>	LATIN SMALL LETTER C WITH COMBINING BREVE	č	U+0043	C
<U+0063,U+0308>	LATIN SMALL LETTER C WITH COMBINING DIAERESIS	ć	U+0043	C
<U+0063,U+0315>	LATIN SMALL LETTER C WITH COMBINING COMMA ABOVE RIGHT	ć	U+0043	C
<U+0063,U+0323>	LATIN SMALL LETTER C WITH COMBINING DOT BELOW	ḥ	U+0043	C
<U+0063,U+0326>	LATIN SMALL LETTER C WITH COMBINING COMMA BELOW	ḥ	U+0043	C
<U+0063,U+0328,U+0306>	LATIN SMALL LETTER C WITH COMBINING OGONEK AND COMBINING BREVE	č	U+0043	C
U+0064	LATIN SMALL LETTER D	d	U+0044	D
<U+0064,U+0302>	LATIN SMALL LETTER D WITH COMBINING CIRCUMFLEX ACCENT	Ď	U+0044	D
U+0065	LATIN SMALL LETTER E	e	U+0045	E
U+0066	LATIN SMALL LETTER F	f	U+0046	F
<U+0066,U+0300>	LATIN SMALL LETTER F WITH COMBINING GRAVE ACCENT	ḑ	U+0046	F
<U+0066,U+0304>	LATIN SMALL LETTER F WITH COMBINING MACRON	ḑ	U+0046	F
U+0067	LATIN SMALL LETTER G	g	U+0047	G
<U+0067,U+0300>	LATIN SMALL LETTER G WITH COMBINING GRAVE ACCENT	ğ	U+0047	G
U+0068	LATIN SMALL LETTER H	h	U+0048	H
<U+0068,U+0304>	LATIN SMALL LETTER H WITH COMBINING MACRON	ĥ	U+0048	H
<U+0068,U+0326>	LATIN SMALL LETTER H WITH COMBINING COMMA BELOW	ḥ	U+0048	H
U+0069	LATIN SMALL LETTER I	i	U+0049	I
U+006A	LATIN SMALL LETTER J	j	U+004A	J
<U+006A,U+0301>	LATIN SMALL LETTER J WITH COMBINING ACUTE ACCENT	ĵ	U+004A	J
U+006B	LATIN SMALL LETTER K	k	U+004B	K
<U+006B,U+0300>	LATIN SMALL LETTER K WITH COMBINING GRAVE ACCENT	ķ	U+004B	K

Quelle			Ziel	
<U+006B,U+0302>	LATIN SMALL LETTER K WITH COMBINING CIRCUMFLEX ACCENT	ķ	U+004B	K
<U+006B,U+0304>	LATIN SMALL LETTER K WITH COMBINING MACRON	ķ̄	U+004B	K
<U+006B,U+0307>	LATIN SMALL LETTER K WITH COMBINING DOT ABOVE	ķ̇	U+004B	K
<U+006B,U+0315>	LATIN SMALL LETTER K WITH COMBINING COMMA ABOVE RIGHT	ķ′	U+004B	K
<U+006B,U+031B>	LATIN SMALL LETTER K WITH COMBINING HORN	ķ̊	U+004B	K
<U+006B,U+0326>	LATIN SMALL LETTER K WITH COMBINING COMMA BELOW	ķ̸	U+004B	K
<U+006B,U+035F,U+0068>	LATIN SMALL LETTER K WITH COMBINING DOUBLE MACRON BELOW AND LATIN SMALL LETTER H	ķ̹	<U+004B,U+0048>	KH
U+006C	LATIN SMALL LETTER L	l	U+004C	L
<U+006C,U+0302>	LATIN SMALL LETTER L WITH COMBINING CIRCUMFLEX ACCENT	ĥ	U+004C	L
<U+006C,U+0325>	LATIN SMALL LETTER L WITH COMBINING RING BELOW	ł	U+004C	L
<U+006C,U+0325,U+0304>	LATIN SMALL LETTER L WITH COMBINING RING BELOW AND COMBINING MACRON	Ł	U+004C	L
<U+006C,U+0326>	LATIN SMALL LETTER L WITH COMBINING COMMA BELOW	ł̸	U+004C	L
U+006D	LATIN SMALL LETTER M	m	U+004D	M
<U+006D,U+0300>	LATIN SMALL LETTER M WITH COMBINING GRAVE ACCENT	ṁ	U+004D	M
<U+006D,U+0302>	LATIN SMALL LETTER M WITH COMBINING CIRCUMFLEX ACCENT	ĥ	U+004D	M
<U+006D,U+0306>	LATIN SMALL LETTER M WITH COMBINING BREVE	ṃ	U+004D	M
<U+006D,U+0310>	LATIN SMALL LETTER M WITH COMBINING CANDRABINDU	Ṅ	U+004D	M
U+006E	LATIN SMALL LETTER N	n	U+004E	N
<U+006E,U+0302>	LATIN SMALL LETTER N WITH COMBINING CIRCUMFLEX ACCENT	ņ	U+004E	N
<U+006E,U+0304>	LATIN SMALL LETTER N WITH COMBINING MACRON	ñ	U+004E	N
<U+006E,U+0306>	LATIN SMALL LETTER N WITH COMBINING BREVE	ṅ	U+004E	N
<U+006E,U+0326>	LATIN SMALL LETTER N WITH COMBINING COMMA BELOW	ṇ̸	U+004E	N
U+006F	LATIN SMALL LETTER O	o	U+004F	O
U+0070	LATIN SMALL LETTER P	p	U+0050	P

Quelle			Ziel	
<U+0070,U+0300>	LATIN SMALL LETTER P WITH COMBINING GRAVE ACCENT	ṑ	U+0050	P
<U+0070,U+0304>	LATIN SMALL LETTER P WITH COMBINING MACRON	ṕ	U+0050	P
<U+0070,U+0315>	LATIN SMALL LETTER P WITH COMBINING COMMA ABOVE RIGHT	ṑ́	U+0050	P
<U+0070,U+0323>	LATIN SMALL LETTER P WITH COMBINING DOT BELOW	ọ̄̀	U+0050	P
U+0071	LATIN SMALL LETTER Q	q	U+0051	Q
U+0072	LATIN SMALL LETTER R	r	U+0052	R
<U+0072,U+0306>	LATIN SMALL LETTER R WITH COMBINING BREVE	ṛ	U+0052	R
<U+0072,U+0325>	LATIN SMALL LETTER R WITH COMBINING RING BELOW	ṛ̣	U+0052	R
<U+0072,U+0325,U+0304>	LATIN SMALL LETTER R WITH COMBINING RING BELOW AND COMBINING MACRON	ṝ̣	U+0052	R
U+0073	LATIN SMALL LETTER S	s	U+0053	S
<U+0073,U+0300>	LATIN SMALL LETTER S WITH COMBINING GRAVE ACCENT	ṣ	U+0053	S
<U+0073,U+0304>	LATIN SMALL LETTER S WITH COMBINING MACRON	ṡ	U+0053	S
<U+0073,U+031B,U+0304>	LATIN SMALL LETTER S WITH COMBINING HORN AND COMBINING MACRON	ṣ̄	U+0053	S
<U+0073,U+0331>	LATIN SMALL LETTER S WITH COMBINING MACRON BELOW	ṩ	U+0053	S
U+0074	LATIN SMALL LETTER T	t	U+0054	T
<U+0074,U+0300>	LATIN SMALL LETTER T WITH COMBINING GRAVE ACCENT	ṭ	U+0054	T
<U+0074,U+0304>	LATIN SMALL LETTER T WITH COMBINING MACRON	ṡ	U+0054	T
<U+0074,U+0315>	LATIN SMALL LETTER T WITH COMBINING COMMA ABOVE RIGHT	ṭ́	U+0054	T
<U+0074,U+031B>	LATIN SMALL LETTER T WITH COMBINING HORN	ṩ	U+0054	T
U+0075	LATIN SMALL LETTER U	u	U+0055	U
<U+0075,U+0307>	LATIN SMALL LETTER U WITH COMBINING DOT ABOVE	ṹ	U+0055	U
U+0076	LATIN SMALL LETTER V	v	U+0056	V
U+0077	LATIN SMALL LETTER W	w	U+0057	W
U+0078	LATIN SMALL LETTER X	x	U+0058	X
U+0079	LATIN SMALL LETTER Y	y	U+0059	Y
U+007A	LATIN SMALL LETTER Z	z	U+005A	Z
<U+007A,U+0300>	LATIN SMALL LETTER Z WITH COMBINING GRAVE ACCENT	ẓ	U+005A	Z

Quelle			Ziel	
<U+007A,U+0304>	LATIN SMALL LETTER Z WITH COMBINING MACRON	z̄	U+005A	Z
<U+007A,U+0306>	LATIN SMALL LETTER Z WITH COMBINING BREVE	z̆	U+005A	Z
<U+007A,U+0308>	LATIN SMALL LETTER Z WITH COMBINING DIAERESIS	z̈	U+005A	Z
<U+007A,U+0327>	LATIN SMALL LETTER Z WITH COMBINING CEDILLA	z̧	U+005A	Z
U+00C0	LATIN CAPITAL LETTER A WITH GRAVE	À	U+0041	A
U+00C1	LATIN CAPITAL LETTER A WITH ACUTE	Á	U+0041	A
U+00C2	LATIN CAPITAL LETTER A WITH CIRCUMFLEX	Â	U+0041	A
U+00C3	LATIN CAPITAL LETTER A WITH TILDE	Ã	U+0041	A
U+00C4	LATIN CAPITAL LETTER A WITH DIAERESIS	Ä	<U+0041,U+0045>	AE
U+00C5	LATIN CAPITAL LETTER A WITH RING ABOVE	Å	<U+0041,U+0041>	AA
U+00C6	LATIN CAPITAL LETTER AE	Æ	<U+0041,U+0045>	AE
U+00C7	LATIN CAPITAL LETTER C WITH CEDILLA	Ç	U+0043	C
<U+00C7,U+0306>	LATIN CAPITAL LETTER C WITH CEDILLA WITH COMBINING BREVE	Ç̆	U+0043	C
U+00C8	LATIN CAPITAL LETTER E WITH GRAVE	È	U+0045	E
U+00C9	LATIN CAPITAL LETTER E WITH ACUTE	É	U+0045	E
U+00CA	LATIN CAPITAL LETTER E WITH CIRCUMFLEX	Ê	U+0045	E
U+00CB	LATIN CAPITAL LETTER E WITH DIAERESIS	Ë	U+0045	E
U+00CC	LATIN CAPITAL LETTER I WITH GRAVE	Ì	U+0049	I
U+00CD	LATIN CAPITAL LETTER I WITH ACUTE	Í	U+0049	I
U+00CE	LATIN CAPITAL LETTER I WITH CIRCUMFLEX	Î	U+0049	I
U+00CF	LATIN CAPITAL LETTER I WITH DIAERESIS	Ï	U+0049	I
U+00D0	LATIN CAPITAL LETTER ETH	Ð	U+0044	D
U+00D1	LATIN CAPITAL LETTER N WITH TILDE	Ñ	U+004E	N
U+00D2	LATIN CAPITAL LETTER O WITH GRAVE	Ò	U+004F	O
U+00D3	LATIN CAPITAL LETTER O WITH ACUTE	Ó	U+004F	O
U+00D4	LATIN CAPITAL LETTER O WITH CIRCUMFLEX	Ô	U+004F	O
U+00D5	LATIN CAPITAL LETTER O WITH TILDE	Õ	U+004F	O
U+00D6	LATIN CAPITAL LETTER O WITH DIAERESIS	Ö	<U+004F,U+0045>	OE
U+00D8	LATIN CAPITAL LETTER O WITH STROKE	Ø	<U+004F,U+0045>	OE
U+00D9	LATIN CAPITAL LETTER U WITH GRAVE	Ù	U+0055	U
U+00DA	LATIN CAPITAL LETTER U WITH ACUTE	Ú	U+0055	U
U+00DB	LATIN CAPITAL LETTER U WITH CIRCUMFLEX	Û	U+0055	U
<U+00DB,U+0304>	LATIN CAPITAL LETTER U WITH CIRCUMFLEX WITH COMBINING MACRON	Û̄	U+0055	U
U+00DC	LATIN CAPITAL LETTER U WITH DIAERESIS	Ü	<U+0055,U+0045>	UE
U+00DD	LATIN CAPITAL LETTER Y WITH ACUTE	Ý	U+0059	Y

Quelle			Ziel	
U+00DE	LATIN CAPITAL LETTER THORN	þ	<U+0054,U+0048>	TH
U+00DF	LATIN SMALL LETTER SHARP S	ß	<U+0053,U+0053>	SS
U+00E0	LATIN SMALL LETTER A WITH GRAVE	à	U+0041	A
U+00E1	LATIN SMALL LETTER A WITH ACUTE	á	U+0041	A
U+00E2	LATIN SMALL LETTER A WITH CIRCUMFLEX	â	U+0041	A
U+00E3	LATIN SMALL LETTER A WITH TILDE	ã	U+0041	A
U+00E4	LATIN SMALL LETTER A WITH DIAERESIS	ä	<U+0041,U+0045>	AE
U+00E5	LATIN SMALL LETTER A WITH RING ABOVE	å	<U+0041,U+0041>	AA
U+00E6	LATIN SMALL LETTER AE	æ	<U+0041,U+0045>	AE
U+00E7	LATIN SMALL LETTER C WITH CEDILLA	ç	U+0043	C
<U+00E7,U+0306>	LATIN SMALL LETTER C WITH CEDILLA WITH COMBINING BREVE	ç̆	U+0043	C
U+00E8	LATIN SMALL LETTER E WITH GRAVE	è	U+0045	E
U+00E9	LATIN SMALL LETTER E WITH ACUTE	é	U+0045	E
U+00EA	LATIN SMALL LETTER E WITH CIRCUMFLEX	ê	U+0045	E
U+00EB	LATIN SMALL LETTER E WITH DIAERESIS	ë	U+0045	E
U+00EC	LATIN SMALL LETTER I WITH GRAVE	ì	U+0049	I
U+00ED	LATIN SMALL LETTER I WITH ACUTE	í	U+0049	I
U+00EE	LATIN SMALL LETTER I WITH CIRCUMFLEX	î	U+0049	I
U+00EF	LATIN SMALL LETTER I WITH DIAERESIS	ï	U+0049	I
U+00F0	LATIN SMALL LETTER ETH	ð	U+0044	D
U+00F1	LATIN SMALL LETTER N WITH TILDE	ñ	U+004E	N
U+00F2	LATIN SMALL LETTER O WITH GRAVE	ò	U+004F	O
U+00F3	LATIN SMALL LETTER O WITH ACUTE	ó	U+004F	O
U+00F4	LATIN SMALL LETTER O WITH CIRCUMFLEX	ô	U+004F	O
U+00F5	LATIN SMALL LETTER O WITH TILDE	õ	U+004F	O
U+00F6	LATIN SMALL LETTER O WITH DIAERESIS	ö	<U+004F,U+0045>	OE
U+00F8	LATIN SMALL LETTER O WITH STROKE	ø	<U+004F,U+0045>	OE
U+00F9	LATIN SMALL LETTER U WITH GRAVE	ù	U+0055	U
U+00FA	LATIN SMALL LETTER U WITH ACUTE	ú	U+0055	U
U+00FB	LATIN SMALL LETTER U WITH CIRCUMFLEX	û	U+0055	U
<U+00FB,U+0304>	LATIN SMALL LETTER U WITH CIRCUMFLEX WITH COMBINING MACRON	ũ	U+0055	U
U+00FC	LATIN SMALL LETTER U WITH DIAERESIS	ü	<U+0055,U+0045>	UE
U+00FD	LATIN SMALL LETTER Y WITH ACUTE	ý	U+0059	Y
U+00FE	LATIN SMALL LETTER THORN	þ	<U+0054,U+0048>	TH
U+00FF	LATIN SMALL LETTER Y WITH DIAERESIS	ÿ	U+0059	Y
<U+00FF,U+0301>	LATIN SMALL LETTER Y WITH DIAERESIS WITH COMBINING ACUTE ACCENT	ÿ̇	U+0059	Y
U+0100	LATIN CAPITAL LETTER A WITH MACRON	Ā	U+0041	A

Quelle			Ziel	
U+0101	LATIN SMALL LETTER A WITH MACRON	ā	U+0041	A
U+0102	LATIN CAPITAL LETTER A WITH BREVE	Ă	U+0041	A
U+0103	LATIN SMALL LETTER A WITH BREVE	ă	U+0041	A
U+0104	LATIN CAPITAL LETTER A WITH OGONEK	Ą	U+0041	A
U+0105	LATIN SMALL LETTER A WITH OGONEK	ą	U+0041	A
U+0106	LATIN CAPITAL LETTER C WITH ACUTE	Ć	U+0043	C
U+0107	LATIN SMALL LETTER C WITH ACUTE	ć	U+0043	C
U+0108	LATIN CAPITAL LETTER C WITH CIRCUMFLEX	Ĉ	U+0043	C
U+0109	LATIN SMALL LETTER C WITH CIRCUMFLEX	ĉ	U+0043	C
U+010A	LATIN CAPITAL LETTER C WITH DOT ABOVE	Ċ	U+0043	C
U+010B	LATIN SMALL LETTER C WITH DOT ABOVE	ċ	U+0043	C
U+010C	LATIN CAPITAL LETTER C WITH CARON	Č	U+0043	C
<U+010C,U+0315>	LATIN CAPITAL LETTER C WITH CARON WITH COMBINING COMMA ABOVE RIGHT	Č̣	U+0043	C
<U+010C,U+0323>	LATIN CAPITAL LETTER C WITH CARON WITH COMBINING DOT BELOW	Č̣̇	U+0043	C
U+010D	LATIN SMALL LETTER C WITH CARON	č	U+0043	C
<U+010D,U+0315>	LATIN SMALL LETTER C WITH CARON WITH COMBINING COMMA ABOVE RIGHT	č̣	U+0043	C
<U+010D,U+0323>	LATIN SMALL LETTER C WITH CARON WITH COMBINING DOT BELOW	č̣̇	U+0043	C
U+010E	LATIN CAPITAL LETTER D WITH CARON	Ď	U+0044	D
U+010F	LATIN SMALL LETTER D WITH CARON	ď	U+0044	D
U+0110	LATIN CAPITAL LETTER D WITH STROKE	Đ	U+0044	D
U+0111	LATIN SMALL LETTER D WITH STROKE	đ	U+0044	D
U+0112	LATIN CAPITAL LETTER E WITH MACRON	Ē	U+0045	E
U+0113	LATIN SMALL LETTER E WITH MACRON	ē	U+0045	E
U+0114	LATIN CAPITAL LETTER E WITH BREVE	Ĕ	U+0045	E
U+0115	LATIN SMALL LETTER E WITH BREVE	ĕ	U+0045	E
U+0116	LATIN CAPITAL LETTER E WITH DOT ABOVE	Ė	U+0045	E
U+0117	LATIN SMALL LETTER E WITH DOT ABOVE	ė	U+0045	E
U+0118	LATIN CAPITAL LETTER E WITH OGONEK	Ę	U+0045	E
U+0119	LATIN SMALL LETTER E WITH OGONEK	ę	U+0045	E
U+011A	LATIN CAPITAL LETTER E WITH CARON	Ě	U+0045	E
U+011B	LATIN SMALL LETTER E WITH CARON	ě	U+0045	E
U+011C	LATIN CAPITAL LETTER G WITH CIRCUMFLEX	Ĝ	U+0047	G
U+011D	LATIN SMALL LETTER G WITH CIRCUMFLEX	ĝ	U+0047	G
U+011E	LATIN CAPITAL LETTER G WITH BREVE	Ğ	U+0047	G
U+011F	LATIN SMALL LETTER G WITH BREVE	ğ	U+0047	G
U+0120	LATIN CAPITAL LETTER G WITH DOT ABOVE	Ġ	U+0047	G

Quelle			Ziel	
U+0121	LATIN SMALL LETTER G WITH DOT ABOVE	ġ	U+0047	G
U+0122	LATIN CAPITAL LETTER G WITH CEDILLA	Ġ	U+0047	G
U+0123	LATIN SMALL LETTER G WITH CEDILLA	ġ	U+0047	G
U+0124	LATIN CAPITAL LETTER H WITH CIRCUMFLEX	Ĥ	U+0048	H
U+0125	LATIN SMALL LETTER H WITH CIRCUMFLEX	ĥ	U+0048	H
U+0126	LATIN CAPITAL LETTER H WITH STROKE	Ħ	U+0048	H
U+0127	LATIN SMALL LETTER H WITH STROKE	ħ	U+0048	H
U+0128	LATIN CAPITAL LETTER I WITH TILDE	Ĩ	U+0049	I
U+0129	LATIN SMALL LETTER I WITH TILDE	ĩ	U+0049	I
U+012A	LATIN CAPITAL LETTER I WITH MACRON	Ī	U+0049	I
<U+012A,U+0301>	LATIN CAPITAL LETTER I WITH MACRON WITH COMBINING ACUTE ACCENT	Ī́	U+0049	I
U+012B	LATIN SMALL LETTER I WITH MACRON	ī	U+0049	I
<U+012B,U+0301>	LATIN SMALL LETTER I WITH MACRON WITH COMBINING ACUTE ACCENT	ī́	U+0049	I
U+012C	LATIN CAPITAL LETTER I WITH BREVE	İ	U+0049	I
U+012D	LATIN SMALL LETTER I WITH BREVE	ı	U+0049	I
U+012E	LATIN CAPITAL LETTER I WITH OGONEK	Į	U+0049	I
U+012F	LATIN SMALL LETTER I WITH OGONEK	į	U+0049	I
U+0130	LATIN CAPITAL LETTER I WITH DOT ABOVE	İ	U+0049	I
U+0131	LATIN SMALL LETTER DOTLESS I	ı	U+0049	I
U+0132	LATIN CAPITAL LIGATURE IJ	IJ	<U+0049,U+004A>	IJ
U+0133	LATIN SMALL LIGATURE IJ	ij	<U+0049,U+004A>	IJ
U+0134	LATIN CAPITAL LETTER J WITH CIRCUMFLEX	Ĵ	U+004A	J
U+0135	LATIN SMALL LETTER J WITH CIRCUMFLEX	ĵ	U+004A	J
U+0136	LATIN CAPITAL LETTER K WITH CEDILLA	Ķ	U+004B	K
U+0137	LATIN SMALL LETTER K WITH CEDILLA	ķ	U+004B	K
U+0138	LATIN SMALL LETTER KRA	κ	U+004B	K
U+0139	LATIN CAPITAL LETTER L WITH ACUTE	Ĺ	U+004C	L
U+013A	LATIN SMALL LETTER L WITH ACUTE	ĺ	U+004C	L
U+013B	LATIN CAPITAL LETTER L WITH CEDILLA	Ľ	U+004C	L
U+013C	LATIN SMALL LETTER L WITH CEDILLA	ļ	U+004C	L
U+013D	LATIN CAPITAL LETTER L WITH CARON	Ľ	U+004C	L
U+013E	LATIN SMALL LETTER L WITH CARON	ľ	U+004C	L
U+013F	LATIN CAPITAL LETTER L WITH MIDDLE DOT	Ľ	U+004C	L
U+0140	LATIN SMALL LETTER L WITH MIDDLE DOT	ḷ	U+004C	L
U+0141	LATIN CAPITAL LETTER L WITH STROKE	Ł	U+004C	L
U+0142	LATIN SMALL LETTER L WITH STROKE	ł	U+004C	L
U+0143	LATIN CAPITAL LETTER N WITH ACUTE	Ñ	U+004E	N
U+0144	LATIN SMALL LETTER N WITH ACUTE	ñ	U+004E	N

Quelle			Ziel	
U+0145	LATIN CAPITAL LETTER N WITH CEDILLA	Ñ	U+004E	N
U+0146	LATIN SMALL LETTER N WITH CEDILLA	ñ	U+004E	N
U+0147	LATIN CAPITAL LETTER N WITH CARON	Ň	U+004E	N
U+0148	LATIN SMALL LETTER N WITH CARON	ň	U+004E	N
U+0149	LATIN SMALL LETTER N PRECEDED BY APOSTROPHE	’n	U+004E	N
U+014A	LATIN CAPITAL LETTER ENG	Ŋ	U+004E	N
U+014B	LATIN SMALL LETTER ENG	ŋ	U+004E	N
U+014C	LATIN CAPITAL LETTER O WITH MACRON	Ō	U+004F	O
U+014D	LATIN SMALL LETTER O WITH MACRON	ō	U+004F	O
U+014E	LATIN CAPITAL LETTER O WITH BREVE	Ŏ	U+004F	O
U+014F	LATIN SMALL LETTER O WITH BREVE	ö	U+004F	O
U+0150	LATIN CAPITAL LETTER O WITH DOUBLE ACUTE	Ő	U+004F	O
U+0151	LATIN SMALL LETTER O WITH DOUBLE ACUTE	ó	U+004F	O
U+0152	LATIN CAPITAL LIGATURE OE	Œ	<U+004F,U+0045>	OE
U+0153	LATIN SMALL LIGATURE OE	œ	<U+004F,U+0045>	OE
U+0154	LATIN CAPITAL LETTER R WITH ACUTE	Ŕ	U+0052	R
U+0155	LATIN SMALL LETTER R WITH ACUTE	ř	U+0052	R
U+0156	LATIN CAPITAL LETTER R WITH CEDILLA	Ŗ	U+0052	R
U+0157	LATIN SMALL LETTER R WITH CEDILLA	ŗ	U+0052	R
U+0158	LATIN CAPITAL LETTER R WITH CARON	Ř	U+0052	R
U+0159	LATIN SMALL LETTER R WITH CARON	ř	U+0052	R
U+015A	LATIN CAPITAL LETTER S WITH ACUTE	Ś	U+0053	S
U+015B	LATIN SMALL LETTER S WITH ACUTE	ś	U+0053	S
U+015C	LATIN CAPITAL LETTER S WITH CIRCUMFLEX	Ŝ	U+0053	S
U+015D	LATIN SMALL LETTER S WITH CIRCUMFLEX	ŝ	U+0053	S
U+015E	LATIN CAPITAL LETTER S WITH CEDILLA	Ş	U+0053	S
U+015F	LATIN SMALL LETTER S WITH CEDILLA	ş	U+0053	S
U+0160	LATIN CAPITAL LETTER S WITH CARON	Š	U+0053	S
U+0161	LATIN SMALL LETTER S WITH CARON	š	U+0053	S
U+0162	LATIN CAPITAL LETTER T WITH CEDILLA	Ţ	U+0054	T
U+0163	LATIN SMALL LETTER T WITH CEDILLA	ţ	U+0054	T
U+0164	LATIN CAPITAL LETTER T WITH CARON	Ť	U+0054	T
U+0165	LATIN SMALL LETTER T WITH CARON	ť	U+0054	T
U+0166	LATIN CAPITAL LETTER T WITH STROKE	Ɽ	U+0054	T
U+0167	LATIN SMALL LETTER T WITH STROKE	ⱥ	U+0054	T
U+0168	LATIN CAPITAL LETTER U WITH TILDE	Ŭ	U+0055	U
U+0169	LATIN SMALL LETTER U WITH TILDE	ũ	U+0055	U
U+016A	LATIN CAPITAL LETTER U WITH MACRON	Ū	U+0055	U
U+016B	LATIN SMALL LETTER U WITH MACRON	ū	U+0055	U

Quelle			Ziel	
U+016C	LATIN CAPITAL LETTER U WITH BREVE	Ů	U+0055	U
U+016D	LATIN SMALL LETTER U WITH BREVE	ů	U+0055	U
U+016E	LATIN CAPITAL LETTER U WITH RING ABOVE	Ű	U+0055	U
U+016F	LATIN SMALL LETTER U WITH RING ABOVE	ű	U+0055	U
U+0170	LATIN CAPITAL LETTER U WITH DOUBLE ACUTE	Ű	U+0055	U
U+0171	LATIN SMALL LETTER U WITH DOUBLE ACUTE	ű	U+0055	U
U+0172	LATIN CAPITAL LETTER U WITH OGONEK	Ų	U+0055	U
U+0173	LATIN SMALL LETTER U WITH OGONEK	ų	U+0055	U
U+0174	LATIN CAPITAL LETTER W WITH CIRCUMFLEX	Ŵ	U+0057	W
U+0175	LATIN SMALL LETTER W WITH CIRCUMFLEX	ŵ	U+0057	W
U+0176	LATIN CAPITAL LETTER Y WITH CIRCUMFLEX	Ŷ	U+0059	Y
U+0177	LATIN SMALL LETTER Y WITH CIRCUMFLEX	ŷ	U+0059	Y
U+0178	LATIN CAPITAL LETTER Y WITH DIAERESIS	ÿ	U+0059	Y
U+0179	LATIN CAPITAL LETTER Z WITH ACUTE	Ż	U+005A	Z
U+017A	LATIN SMALL LETTER Z WITH ACUTE	ź	U+005A	Z
U+017B	LATIN CAPITAL LETTER Z WITH DOT ABOVE	Ž	U+005A	Z
U+017C	LATIN SMALL LETTER Z WITH DOT ABOVE	ž	U+005A	Z
U+017D	LATIN CAPITAL LETTER Z WITH CARON	Ž	U+005A	Z
<U+017D,U+0326>	LATIN CAPITAL LETTER Z WITH CARON WITH COMBINING COMMA BELOW	Ẓ̌	U+005A	Z
<U+017D,U+0327>	LATIN CAPITAL LETTER Z WITH CARON WITH COMBINING CEDILLA	Ž̧	U+005A	Z
U+017E	LATIN SMALL LETTER Z WITH CARON	ž	U+005A	Z
<U+017E,U+0326>	LATIN SMALL LETTER Z WITH CARON WITH COMBINING COMMA BELOW	ẓ̌	U+005A	Z
<U+017E,U+0327>	LATIN SMALL LETTER Z WITH CARON WITH COMBINING CEDILLA	ž̧	U+005A	Z
U+0187	LATIN CAPITAL LETTER C WITH HOOK	Č	U+0043	C
U+0188	LATIN SMALL LETTER C WITH HOOK	č	U+0043	C
U+018F	LATIN CAPITAL LETTER SCHWA	Ə	U+0045	E
U+0197	LATIN CAPITAL LETTER I WITH STROKE	Ɔ	U+0049	I
U+01A0	LATIN CAPITAL LETTER O WITH HORN	Ɔ	U+004F	O
U+01A1	LATIN SMALL LETTER O WITH HORN	ɔ	U+004F	O
U+01AF	LATIN CAPITAL LETTER U WITH HORN	Ɔ	U+0055	U
U+01B0	LATIN SMALL LETTER U WITH HORN	ɔ	U+0055	U
U+01B7	LATIN CAPITAL LETTER EZH	Ɔ	U+005A	Z
U+01CD	LATIN CAPITAL LETTER A WITH CARON	Ǻ	U+0041	A
U+01CE	LATIN SMALL LETTER A WITH CARON	ǻ	U+0041	A
U+01CF	LATIN CAPITAL LETTER I WITH CARON	Ǫ	U+0049	I
U+01D0	LATIN SMALL LETTER I WITH CARON	ǫ	U+0049	I

Quelle			Ziel	
U+01D1	LATIN CAPITAL LETTER O WITH CARON	Ŏ	U+004F	O
U+01D2	LATIN SMALL LETTER O WITH CARON	ő	U+004F	O
U+01D3	LATIN CAPITAL LETTER U WITH CARON	Ů	U+0055	U
U+01D4	LATIN SMALL LETTER U WITH CARON	ů	U+0055	U
U+01D5	LATIN CAPITAL LETTER U WITH DIAERESIS AND MACRON	Ū	<U+0055,U+0045>	UE
U+01D6	LATIN SMALL LETTER U WITH DIAERESIS AND MACRON	ū	<U+0055,U+0045>	UE
U+01D7	LATIN CAPITAL LETTER U WITH DIAERESIS AND ACUTE	Ű	<U+0055,U+0045>	UE
U+01D8	LATIN SMALL LETTER U WITH DIAERESIS AND ACUTE	ű	<U+0055,U+0045>	UE
U+01D9	LATIN CAPITAL LETTER U WITH DIAERESIS AND CARON	Ŭ	<U+0055,U+0045>	UE
U+01DA	LATIN SMALL LETTER U WITH DIAERESIS AND CARON	ű	<U+0055,U+0045>	UE
U+01DB	LATIN CAPITAL LETTER U WITH DIAERESIS AND GRAVE	Ŭ	<U+0055,U+0045>	UE
U+01DC	LATIN SMALL LETTER U WITH DIAERESIS AND GRAVE	ű	<U+0055,U+0045>	UE
U+01DE	LATIN CAPITAL LETTER A WITH DIAERESIS AND MACRON	Ä	<U+0041,U+0045>	AE
U+01DF	LATIN SMALL LETTER A WITH DIAERESIS AND MACRON	ä	<U+0041,U+0045>	AE
U+01E2	LATIN CAPITAL LETTER AE WITH MACRON	Ā	<U+0041,U+0045>	AE
U+01E3	LATIN SMALL LETTER AE WITH MACRON	æ	<U+0041,U+0045>	AE
U+01E4	LATIN CAPITAL LETTER G WITH STROKE	G	U+0047	G
U+01E5	LATIN SMALL LETTER G WITH STROKE	g	U+0047	G
U+01E6	LATIN CAPITAL LETTER G WITH CARON	Ĝ	U+0047	G
U+01E7	LATIN SMALL LETTER G WITH CARON	ĝ	U+0047	G
U+01E8	LATIN CAPITAL LETTER K WITH CARON	Ķ	U+004B	K
U+01E9	LATIN SMALL LETTER K WITH CARON	ķ	U+004B	K
U+01EA	LATIN CAPITAL LETTER O WITH OGONEK	Q	U+004F	O
U+01EB	LATIN SMALL LETTER O WITH OGONEK	q	U+004F	O
U+01EC	LATIN CAPITAL LETTER O WITH OGONEK AND MACRON	Ų	U+004F	O
U+01ED	LATIN SMALL LETTER O WITH OGONEK AND MACRON	ō	U+004F	O
U+01EE	LATIN CAPITAL LETTER EZH WITH CARON	Š	U+005A	Z
U+01EF	LATIN SMALL LETTER EZH WITH CARON	š	U+005A	Z
U+01F0	LATIN SMALL LETTER J WITH CARON	Ĵ	U+004A	J
U+01F4	LATIN CAPITAL LETTER G WITH ACUTE	Ġ	U+0047	G

Quelle			Ziel	
U+01F5	LATIN SMALL LETTER G WITH ACUTE	ǧ	U+0047	G
U+01F8	LATIN CAPITAL LETTER N WITH GRAVE	Ñ	U+004E	N
U+01F9	LATIN SMALL LETTER N WITH GRAVE	ñ	U+004E	N
U+01FA	LATIN CAPITAL LETTER A WITH RING ABOVE AND ACUTE	Ǻ	<U+0041,U+0041>	AA
U+01FB	LATIN SMALL LETTER A WITH RING ABOVE AND ACUTE	ǻ	<U+0041,U+0041>	AA
U+01FC	LATIN CAPITAL LETTER AE WITH ACUTE	Æ	<U+0041,U+0045>	AE
U+01FD	LATIN SMALL LETTER AE WITH ACUTE	æ	<U+0041,U+0045>	AE
U+01FE	LATIN CAPITAL LETTER O WITH STROKE AND ACUTE	Ø	<U+004F,U+0045>	OE
U+01FF	LATIN SMALL LETTER O WITH STROKE AND ACUTE	ø	<U+004F,U+0045>	OE
U+0212	LATIN CAPITAL LETTER R WITH INVERTED BREVE	Ŕ	U+0052	R
U+0213	LATIN SMALL LETTER R WITH INVERTED BREVE	ŕ	U+0052	R
U+0218	LATIN CAPITAL LETTER S WITH COMMA BELOW	Ș	U+0053	S
U+0219	LATIN SMALL LETTER S WITH COMMA BELOW	ș	U+0053	S
U+021A	LATIN CAPITAL LETTER T WITH COMMA BELOW	Ț	U+0054	T
U+021B	LATIN SMALL LETTER T WITH COMMA BELOW	ț	U+0054	T
U+021E	LATIN CAPITAL LETTER H WITH CARON	Ĥ	U+0048	H
U+021F	LATIN SMALL LETTER H WITH CARON	ĥ	U+0048	H
U+0227	LATIN SMALL LETTER A WITH DOT ABOVE	ȁ	U+0041	A
U+0228	LATIN CAPITAL LETTER E WITH CEDILLA	Ẹ	U+0045	E
U+0229	LATIN SMALL LETTER E WITH CEDILLA	ẹ	U+0045	E
U+022A	LATIN CAPITAL LETTER O WITH DIAERESIS AND MACRON	Ȫ	<U+004F,U+0045>	OE
U+022B	LATIN SMALL LETTER O WITH DIAERESIS AND MACRON	ȫ	<U+004F,U+0045>	OE
U+022C	LATIN CAPITAL LETTER O WITH TILDE AND MACRON	Ȫ	U+004F	O
U+022D	LATIN SMALL LETTER O WITH TILDE AND MACRON	ȫ	U+004F	O
U+022E	LATIN CAPITAL LETTER O WITH DOT ABOVE	Ȭ	U+004F	O
U+022F	LATIN SMALL LETTER O WITH DOT ABOVE	ȭ	U+004F	O
U+0230	LATIN CAPITAL LETTER O WITH DOT ABOVE AND MACRON	Ȯ	U+004F	O
U+0231	LATIN SMALL LETTER O WITH DOT ABOVE AND MACRON	ȯ	U+004F	O
U+0232	LATIN CAPITAL LETTER Y WITH MACRON	Ȱ	U+0059	Y
U+0233	LATIN SMALL LETTER Y WITH MACRON	ȱ	U+0059	Y
U+0259	LATIN SMALL LETTER SCHWA	ə	U+0045	E
U+0268	LATIN SMALL LETTER I WITH STROKE	ı	U+0049	I

Quelle			Ziel	
U+0292	LATIN SMALL LETTER EZH	Ʒ	U+005A	Z
U+1E02	LATIN CAPITAL LETTER B WITH DOT ABOVE	Ĭ	U+0042	B
U+1E03	LATIN SMALL LETTER B WITH DOT ABOVE	ĭ	U+0042	B
U+1E06	LATIN CAPITAL LETTER B WITH LINE BELOW	Ḃ	U+0042	B
U+1E07	LATIN SMALL LETTER B WITH LINE BELOW	ḃ	U+0042	B
U+1E0A	LATIN CAPITAL LETTER D WITH DOT ABOVE	Ĭ	U+0044	D
U+1E0B	LATIN SMALL LETTER D WITH DOT ABOVE	ĭ	U+0044	D
U+1E0C	LATIN CAPITAL LETTER D WITH DOT BELOW	Ḑ	U+0044	D
U+1E0D	LATIN SMALL LETTER D WITH DOT BELOW	ḑ	U+0044	D
U+1E0E	LATIN CAPITAL LETTER D WITH LINE BELOW	Ḓ	U+0044	D
U+1E0F	LATIN SMALL LETTER D WITH LINE BELOW	ḓ	U+0044	D
U+1E10	LATIN CAPITAL LETTER D WITH CEDILLA	Ḕ	U+0044	D
U+1E11	LATIN SMALL LETTER D WITH CEDILLA	ḕ	U+0044	D
U+1E1C	LATIN CAPITAL LETTER E WITH CEDILLA AND BREVE	Ḟ	U+0045	E
U+1E1D	LATIN SMALL LETTER E WITH CEDILLA AND BREVE	ḟ	U+0045	E
U+1E1E	LATIN CAPITAL LETTER F WITH DOT ABOVE	Ĭ	U+0046	F
U+1E1F	LATIN SMALL LETTER F WITH DOT ABOVE	ĭ	U+0046	F
U+1E20	LATIN CAPITAL LETTER G WITH MACRON	Ḡ	U+0047	G
U+1E21	LATIN SMALL LETTER G WITH MACRON	ḡ	U+0047	G
U+1E22	LATIN CAPITAL LETTER H WITH DOT ABOVE	Ĭ	U+0048	H
U+1E23	LATIN SMALL LETTER H WITH DOT ABOVE	ĭ	U+0048	H
U+1E24	LATIN CAPITAL LETTER H WITH DOT BELOW	Ḑ	U+0048	H
U+1E25	LATIN SMALL LETTER H WITH DOT BELOW	ḑ	U+0048	H
U+1E26	LATIN CAPITAL LETTER H WITH DIAERESIS	Ḣ	U+0048	H
U+1E27	LATIN SMALL LETTER H WITH DIAERESIS	ḣ	U+0048	H
U+1E28	LATIN CAPITAL LETTER H WITH CEDILLA	Ḥ	U+0048	H
U+1E29	LATIN SMALL LETTER H WITH CEDILLA	ḥ	U+0048	H
U+1E2A	LATIN CAPITAL LETTER H WITH BREVE BELOW	Ḧ	U+0048	H
U+1E2B	LATIN SMALL LETTER H WITH BREVE BELOW	ḧ	U+0048	H
U+1E2F	LATIN SMALL LETTER I WITH DIAERESIS AND ACUTE	Ḱ	U+0049	I
U+1E30	LATIN CAPITAL LETTER K WITH ACUTE	Ḳ	U+004B	K
U+1E31	LATIN SMALL LETTER K WITH ACUTE	ḳ	U+004B	K
U+1E32	LATIN CAPITAL LETTER K WITH DOT BELOW	Ḵ	U+004B	K
<U+1E32,U+0304>	LATIN CAPITAL LETTER K WITH DOT BELOW WITH COMBINING MACRON	Ḵ̄	U+004B	K
U+1E33	LATIN SMALL LETTER K WITH DOT BELOW	ḵ	U+004B	K
<U+1E33,U+0304>	LATIN SMALL LETTER K WITH DOT BELOW WITH COMBINING MACRON	ḵ̄	U+004B	K

Quelle			Ziel	
U+1E34	LATIN CAPITAL LETTER K WITH LINE BELOW	Ƙ	U+004B	K
U+1E35	LATIN SMALL LETTER K WITH LINE BELOW	ƙ	U+004B	K
U+1E36	LATIN CAPITAL LETTER L WITH DOT BELOW	Ɩ	U+004C	L
U+1E37	LATIN SMALL LETTER L WITH DOT BELOW	Ɵ	U+004C	L
U+1E3A	LATIN CAPITAL LETTER L WITH LINE BELOW	Ł	U+004C	L
U+1E3B	LATIN SMALL LETTER L WITH LINE BELOW	ł	U+004C	L
U+1E40	LATIN CAPITAL LETTER M WITH DOT ABOVE	Ṁ	U+004D	M
U+1E41	LATIN SMALL LETTER M WITH DOT ABOVE	ṁ	U+004D	M
U+1E42	LATIN CAPITAL LETTER M WITH DOT BELOW	Ṃ	U+004D	M
U+1E43	LATIN SMALL LETTER M WITH DOT BELOW	ṃ	U+004D	M
U+1E44	LATIN CAPITAL LETTER N WITH DOT ABOVE	Ṇ	U+004E	N
U+1E45	LATIN SMALL LETTER N WITH DOT ABOVE	ṇ	U+004E	N
U+1E46	LATIN CAPITAL LETTER N WITH DOT BELOW	Ṅ	U+004E	N
U+1E47	LATIN SMALL LETTER N WITH DOT BELOW	ṅ	U+004E	N
U+1E48	LATIN CAPITAL LETTER N WITH LINE BELOW	ƚ	U+004E	N
U+1E49	LATIN SMALL LETTER N WITH LINE BELOW	ƞ	U+004E	N
U+1E52	LATIN CAPITAL LETTER O WITH MACRON AND ACUTE	Ȫ	U+004F	O
U+1E53	LATIN SMALL LETTER O WITH MACRON AND ACUTE	ȫ	U+004F	O
U+1E54	LATIN CAPITAL LETTER P WITH ACUTE	Ṕ	U+0050	P
U+1E55	LATIN SMALL LETTER P WITH ACUTE	ṑ	U+0050	P
U+1E56	LATIN CAPITAL LETTER P WITH DOT ABOVE	Ṗ	U+0050	P
U+1E57	LATIN SMALL LETTER P WITH DOT ABOVE	ṗ	U+0050	P
U+1E58	LATIN CAPITAL LETTER R WITH DOT ABOVE	Ṛ	U+0052	R
U+1E59	LATIN SMALL LETTER R WITH DOT ABOVE	ṛ	U+0052	R
U+1E5A	LATIN CAPITAL LETTER R WITH DOT BELOW	Ṙ	U+0052	R
U+1E5B	LATIN SMALL LETTER R WITH DOT BELOW	ṙ	U+0052	R
U+1E5E	LATIN CAPITAL LETTER R WITH LINE BELOW	ƚ	U+0052	R
U+1E5F	LATIN SMALL LETTER R WITH LINE BELOW	ƞ	U+0052	R
U+1E60	LATIN CAPITAL LETTER S WITH DOT ABOVE	Ṡ	U+0053	S
U+1E61	LATIN SMALL LETTER S WITH DOT ABOVE	ṡ	U+0053	S
U+1E62	LATIN CAPITAL LETTER S WITH DOT BELOW	Ṣ	U+0053	S
<U+1E62,U+0304>	LATIN CAPITAL LETTER S WITH DOT BELOW WITH COMBINING MACRON	Ṝ	U+0053	S
U+1E63	LATIN SMALL LETTER S WITH DOT BELOW	ṣ	U+0053	S
<U+1E63,U+0304>	LATIN SMALL LETTER S WITH DOT BELOW WITH COMBINING MACRON	ṝ	U+0053	S
U+1E6A	LATIN CAPITAL LETTER T WITH DOT ABOVE	Ṛ	U+0054	T
U+1E6B	LATIN SMALL LETTER T WITH DOT ABOVE	ṛ	U+0054	T

Quelle			Ziel	
U+1E6C	LATIN CAPITAL LETTER T WITH DOT BELOW	Ṭ	U+0054	T
<U+1E6C,U+0304>	LATIN CAPITAL LETTER T WITH DOT BELOW WITH COMBINING MACRON	Ṯ	U+0054	T
U+1E6D	LATIN SMALL LETTER T WITH DOT BELOW	ṭ	U+0054	T
<U+1E6D,U+0304>	LATIN SMALL LETTER T WITH DOT BELOW WITH COMBINING MACRON	ṯ	U+0054	T
U+1E6E	LATIN CAPITAL LETTER T WITH LINE BELOW	Ṛ	U+0054	T
U+1E6F	LATIN SMALL LETTER T WITH LINE BELOW	ṛ	U+0054	T
U+1E80	LATIN CAPITAL LETTER W WITH GRAVE	Ẁ	U+0057	W
U+1E81	LATIN SMALL LETTER W WITH GRAVE	ẁ	U+0057	W
U+1E82	LATIN CAPITAL LETTER W WITH ACUTE	Ẃ	U+0057	W
U+1E83	LATIN SMALL LETTER W WITH ACUTE	ẃ	U+0057	W
U+1E84	LATIN CAPITAL LETTER W WITH DIAERESIS	Ẅ	U+0057	W
U+1E85	LATIN SMALL LETTER W WITH DIAERESIS	ẅ	U+0057	W
U+1E86	LATIN CAPITAL LETTER W WITH DOT ABOVE	Ẇ	U+0057	W
U+1E87	LATIN SMALL LETTER W WITH DOT ABOVE	ẇ	U+0057	W
U+1E8C	LATIN CAPITAL LETTER X WITH DIAERESIS	Ẋ	U+0058	X
U+1E8D	LATIN SMALL LETTER X WITH DIAERESIS	ẋ	U+0058	X
U+1E8E	LATIN CAPITAL LETTER Y WITH DOT ABOVE	Ỳ	U+0059	Y
U+1E8F	LATIN SMALL LETTER Y WITH DOT ABOVE	ỳ	U+0059	Y
U+1E90	LATIN CAPITAL LETTER Z WITH CIRCUMFLEX	Ẑ	U+005A	Z
U+1E91	LATIN SMALL LETTER Z WITH CIRCUMFLEX	ẑ	U+005A	Z
U+1E92	LATIN CAPITAL LETTER Z WITH DOT BELOW	Ẓ	U+005A	Z
U+1E93	LATIN SMALL LETTER Z WITH DOT BELOW	ẓ	U+005A	Z
U+1E94	LATIN CAPITAL LETTER Z WITH LINE BELOW	Ẕ	U+005A	Z
U+1E95	LATIN SMALL LETTER Z WITH LINE BELOW	ẕ	U+005A	Z
U+1E96	LATIN SMALL LETTER H WITH LINE BELOW	ḣ	U+0048	H
U+1E97	LATIN SMALL LETTER T WITH DIAERESIS	ṫ	U+0054	T
U+1E9E	LATIN CAPITAL LETTER SHARP S	ß	<U+0053,U+0053>	SS
U+1EA0	LATIN CAPITAL LETTER A WITH DOT BELOW	Ḑ	U+0041	A
<U+1EA0,U+0308>	LATIN CAPITAL LETTER A WITH DOT BELOW WITH COMBINING DIAERESIS	Ḓ	<U+0041,U+0045>	AE
U+1EA1	LATIN SMALL LETTER A WITH DOT BELOW	ḑ	U+0041	A
<U+1EA1,U+0308>	LATIN SMALL LETTER A WITH DOT BELOW WITH COMBINING DIAERESIS	ḓ	<U+0041,U+0045>	AE
U+1EA2	LATIN CAPITAL LETTER A WITH HOOK ABOVE	Ḃ	U+0041	A
U+1EA3	LATIN SMALL LETTER A WITH HOOK ABOVE	ḃ	U+0041	A
U+1EA4	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND ACUTE	Ḅ	U+0041	A
U+1EA5	LATIN SMALL LETTER A WITH CIRCUMFLEX AND ACUTE	ḅ	U+0041	A

Quelle			Ziel	
U+1EA6	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND GRAVE	À	U+0041	A
U+1EA7	LATIN SMALL LETTER A WITH CIRCUMFLEX AND GRAVE	à	U+0041	A
U+1EA8	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE	Ā	U+0041	A
U+1EA9	LATIN SMALL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE	ā	U+0041	A
U+1EAA	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND TILDE	Ã	U+0041	A
U+1EAB	LATIN SMALL LETTER A WITH CIRCUMFLEX AND TILDE	ã	U+0041	A
U+1EAC	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND DOT BELOW	Â	U+0041	A
U+1EAD	LATIN SMALL LETTER A WITH CIRCUMFLEX AND DOT BELOW	â	U+0041	A
U+1EAE	LATIN CAPITAL LETTER A WITH BREVE AND ACUTE	Ǻ	U+0041	A
U+1EAF	LATIN SMALL LETTER A WITH BREVE AND ACUTE	ǻ	U+0041	A
U+1EB0	LATIN CAPITAL LETTER A WITH BREVE AND GRAVE	Ǽ	U+0041	A
U+1EB1	LATIN SMALL LETTER A WITH BREVE AND GRAVE	ǽ	U+0041	A
U+1EB2	LATIN CAPITAL LETTER A WITH BREVE AND HOOK ABOVE	Ǿ	U+0041	A
U+1EB3	LATIN SMALL LETTER A WITH BREVE AND HOOK ABOVE	ǿ	U+0041	A
U+1EB4	LATIN CAPITAL LETTER A WITH BREVE AND TILDE	Ǻ	U+0041	A
U+1EB5	LATIN SMALL LETTER A WITH BREVE AND TILDE	ǻ	U+0041	A
U+1EB6	LATIN CAPITAL LETTER A WITH BREVE AND DOT BELOW	Ǽ	U+0041	A
U+1EB7	LATIN SMALL LETTER A WITH BREVE AND DOT BELOW	ǽ	U+0041	A
U+1EB8	LATIN CAPITAL LETTER E WITH DOT BELOW	Ǝ	U+0045	E
U+1EB9	LATIN SMALL LETTER E WITH DOT BELOW	Ǝ	U+0045	E
U+1EBA	LATIN CAPITAL LETTER E WITH HOOK ABOVE	Ě	U+0045	E
U+1EBB	LATIN SMALL LETTER E WITH HOOK ABOVE	ě	U+0045	E
U+1EBC	LATIN CAPITAL LETTER E WITH TILDE	Ễ	U+0045	E
U+1EBD	LATIN SMALL LETTER E WITH TILDE	ẽ	U+0045	E
U+1EBE	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND ACUTE	Ê	U+0045	E

Quelle			Ziel	
U+1EBF	LATIN SMALL LETTER E WITH CIRCUMFLEX AND ACUTE	é	U+0045	E
U+1EC0	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND GRAVE	Ê	U+0045	E
U+1EC1	LATIN SMALL LETTER E WITH CIRCUMFLEX AND GRAVE	è	U+0045	E
U+1EC2	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE	Ễ	U+0045	E
U+1EC3	LATIN SMALL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE	ễ	U+0045	E
U+1EC4	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND TILDE	Ễ	U+0045	E
U+1EC5	LATIN SMALL LETTER E WITH CIRCUMFLEX AND TILDE	ễ	U+0045	E
U+1EC6	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND DOT BELOW	Ệ	U+0045	E
U+1EC7	LATIN SMALL LETTER E WITH CIRCUMFLEX AND DOT BELOW	ệ	U+0045	E
U+1EC8	LATIN CAPITAL LETTER I WITH HOOK ABOVE	İ̇	U+0049	I
U+1EC9	LATIN SMALL LETTER I WITH HOOK ABOVE	ı̇	U+0049	I
U+1ECA	LATIN CAPITAL LETTER I WITH DOT BELOW	Ị̇	U+0049	I
U+1ECB	LATIN SMALL LETTER I WITH DOT BELOW	ı̣	U+0049	I
U+1ECC	LATIN CAPITAL LETTER O WITH DOT BELOW	Ȯ	U+004F	O
<U+1ECC,U+0308>	LATIN CAPITAL LETTER O WITH DOT BELOW WITH COMBINING DIAERESIS	Ȯ̈	<U+004F,U+0045>	OE
U+1ECD	LATIN SMALL LETTER O WITH DOT BELOW	ȯ	U+004F	O
<U+1ECD,U+0308>	LATIN SMALL LETTER O WITH DOT BELOW WITH COMBINING DIAERESIS	ȯ̈	<U+004F,U+0045>	OE
U+1ECE	LATIN CAPITAL LETTER O WITH HOOK ABOVE	Ȫ	U+004F	O
U+1ECF	LATIN SMALL LETTER O WITH HOOK ABOVE	ȫ	U+004F	O
U+1ED0	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND ACUTE	Ô	U+004F	O
U+1ED1	LATIN SMALL LETTER O WITH CIRCUMFLEX AND ACUTE	ô	U+004F	O
U+1ED2	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND GRAVE	Û	U+004F	O
U+1ED3	LATIN SMALL LETTER O WITH CIRCUMFLEX AND GRAVE	õ	U+004F	O
U+1ED4	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE	Û̃	U+004F	O
U+1ED5	LATIN SMALL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE	õ̃	U+004F	O
U+1ED6	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND TILDE	Û̃	U+004F	O

Quelle			Ziel	
U+1ED7	LATIN SMALL LETTER O WITH CIRCUMFLEX AND TILDE	õ	U+004F	O
U+1ED8	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND DOT BELOW	Ộ	U+004F	O
U+1ED9	LATIN SMALL LETTER O WITH CIRCUMFLEX AND DOT BELOW	ộ	U+004F	O
U+1EDA	LATIN CAPITAL LETTER O WITH HORN AND ACUTE	Ớ	U+004F	O
U+1EDB	LATIN SMALL LETTER O WITH HORN AND ACUTE	ớ	U+004F	O
U+1EDC	LATIN CAPITAL LETTER O WITH HORN AND GRAVE	Ỗ	U+004F	O
U+1EDD	LATIN SMALL LETTER O WITH HORN AND GRAVE	ở	U+004F	O
U+1EDE	LATIN CAPITAL LETTER O WITH HORN AND HOOK ABOVE	Ỗ̃	U+004F	O
U+1EDF	LATIN SMALL LETTER O WITH HORN AND HOOK ABOVE	ở̃	U+004F	O
U+1EE0	LATIN CAPITAL LETTER O WITH HORN AND TILDE	Ỗ̃	U+004F	O
U+1EE1	LATIN SMALL LETTER O WITH HORN AND TILDE	ở̃	U+004F	O
U+1EE2	LATIN CAPITAL LETTER O WITH HORN AND DOT BELOW	Ộ̃	U+004F	O
U+1EE3	LATIN SMALL LETTER O WITH HORN AND DOT BELOW	ợ̉	U+004F	O
U+1EE4	LATIN CAPITAL LETTER U WITH DOT BELOW	Ụ	U+0055	U
<U+1EE4,U+0304>	LATIN CAPITAL LETTER U WITH DOT BELOW WITH COMBINING MACRON	Ŭ	U+0055	U
<U+1EE4,U+0308>	LATIN CAPITAL LETTER U WITH DOT BELOW WITH COMBINING DIAERESIS	Ü	<U+0055,U+0045>	UE
U+1EE5	LATIN SMALL LETTER U WITH DOT BELOW	ụ	U+0055	U
<U+1EE5,U+0304>	LATIN SMALL LETTER U WITH DOT BELOW WITH COMBINING MACRON	ŭ	U+0055	U
<U+1EE5,U+0308>	LATIN SMALL LETTER U WITH DOT BELOW WITH COMBINING DIAERESIS	ü	<U+0055,U+0045>	UE
U+1EE6	LATIN CAPITAL LETTER U WITH HOOK ABOVE	Ủ̃	U+0055	U
U+1EE7	LATIN SMALL LETTER U WITH HOOK ABOVE	ủ̃	U+0055	U
U+1EE8	LATIN CAPITAL LETTER U WITH HORN AND ACUTE	Ứ	U+0055	U
U+1EE9	LATIN SMALL LETTER U WITH HORN AND ACUTE	ứ	U+0055	U
U+1EEA	LATIN CAPITAL LETTER U WITH HORN AND GRAVE	Ỡ	U+0055	U
U+1EEB	LATIN SMALL LETTER U WITH HORN AND GRAVE	ở	U+0055	U

Quelle			Ziel	
U+1EEC	LATIN CAPITAL LETTER U WITH HORN AND HOOK ABOVE	Ů̂	U+0055	U
U+1EED	LATIN SMALL LETTER U WITH HORN AND HOOK ABOVE	ů̂	U+0055	U
U+1EEE	LATIN CAPITAL LETTER U WITH HORN AND TILDE	Ů̃	U+0055	U
U+1EEF	LATIN SMALL LETTER U WITH HORN AND TILDE	ů̃	U+0055	U
U+1EF0	LATIN CAPITAL LETTER U WITH HORN AND DOT BELOW	Ụ̊	U+0055	U
U+1EF1	LATIN SMALL LETTER U WITH HORN AND DOT BELOW	ụ̊	U+0055	U
U+1EF2	LATIN CAPITAL LETTER Y WITH GRAVE	Ỳ	U+0059	Y
U+1EF3	LATIN SMALL LETTER Y WITH GRAVE	ỳ	U+0059	Y
U+1EF4	LATIN CAPITAL LETTER Y WITH DOT BELOW	Ỵ̀	U+0059	Y
U+1EF5	LATIN SMALL LETTER Y WITH DOT BELOW	ỵ̀	U+0059	Y
U+1EF6	LATIN CAPITAL LETTER Y WITH HOOK ABOVE	Ỳ̂	U+0059	Y
U+1EF7	LATIN SMALL LETTER Y WITH HOOK ABOVE	ỳ̂	U+0059	Y
U+1EF8	LATIN CAPITAL LETTER Y WITH TILDE	Ỳ̃	U+0059	Y
U+1EF9	LATIN SMALL LETTER Y WITH TILDE	ỳ̃	U+0059	Y

Anhang A. Erweiterte Schriftzeichen

A.1. Allgemeines

Jedes IT-Verfahren sollte die in Tabelle A.1 bis Tabelle A.3 genannten Schriftzeichen bei allen für Namen (im weiteren Sinne) bestimmten Datenfeldern in vollem Umfang verarbeiten können. Dies umfasst insbesondere die Erfassung, Speicherung, Übermittlung, Anzeige und den Ausdruck aller Schriftzeichen.

A.2. Griechische Buchstaben

Jedes IT-Verfahren sollte die in Tabelle A.1 aufgeführten Schriftzeichen unterstützen, die insbesondere für den Datenaustausch griechischer Namen (im weiteren Sinne) im europäischen Kontext benötigt werden. Die griechischen Schriftzeichen sollten unterstützt werden, um für den grenzüberschreitenden Datenaustausch griechische Namen im unveränderten Original übertragen zu können.

Tabelle A.1. Griechische Buchstaben

Codepoints	Name	Glyphe
U+0386	GREEK CAPITAL LETTER ALPHA WITH TONOS	Ά
U+0388	GREEK CAPITAL LETTER EPSILON WITH TONOS	Έ
U+0389	GREEK CAPITAL LETTER ETA WITH TONOS	Ή
U+038A	GREEK CAPITAL LETTER IOTA WITH TONOS	Ί
U+038C	GREEK CAPITAL LETTER OMICRON WITH TONOS	Ό
U+038E	GREEK CAPITAL LETTER UPSILON WITH TONOS	Υ
U+038F	GREEK CAPITAL LETTER OMEGA WITH TONOS	Ω
U+0390	GREEK SMALL LETTER IOTA WITH DIALYTIKA AND TONOS	ϊ
U+0391	GREEK CAPITAL LETTER ALPHA	Α
U+0392	GREEK CAPITAL LETTER BETA	Β
U+0393	GREEK CAPITAL LETTER GAMMA	Γ
U+0394	GREEK CAPITAL LETTER DELTA	Δ
U+0395	GREEK CAPITAL LETTER EPSILON	Ε
U+0396	GREEK CAPITAL LETTER ZETA	Ζ
U+0397	GREEK CAPITAL LETTER ETA	Η
U+0398	GREEK CAPITAL LETTER THETA	Θ
U+0399	GREEK CAPITAL LETTER IOTA	Ι
U+039A	GREEK CAPITAL LETTER KAPPA	Κ
U+039B	GREEK CAPITAL LETTER LAMDA	Λ
U+039C	GREEK CAPITAL LETTER MU	Μ
U+039D	GREEK CAPITAL LETTER NU	Ν
U+039E	GREEK CAPITAL LETTER XI	Ξ
U+039F	GREEK CAPITAL LETTER OMICRON	Ο
U+03A0	GREEK CAPITAL LETTER PI	Π
U+03A1	GREEK CAPITAL LETTER RHO	Ρ

Codepoints	Name	Glyphe
U+03A3	GREEK CAPITAL LETTER SIGMA	Σ
U+03A4	GREEK CAPITAL LETTER TAU	Τ
U+03A5	GREEK CAPITAL LETTER UPSILON	Υ
U+03A6	GREEK CAPITAL LETTER PHI	Φ
U+03A7	GREEK CAPITAL LETTER CHI	Χ
U+03A8	GREEK CAPITAL LETTER PSI	Ψ
U+03A9	GREEK CAPITAL LETTER OMEGA	Ω
U+03AA	GREEK CAPITAL LETTER IOTA WITH DIALYTIKA	ϊ
U+03AB	GREEK CAPITAL LETTER UPSILON WITH DIALYTIKA	ϋ
U+03AC	GREEK SMALL LETTER ALPHA WITH TONOS	ά
U+03AD	GREEK SMALL LETTER EPSILON WITH TONOS	έ
U+03AE	GREEK SMALL LETTER ETA WITH TONOS	ή
U+03AF	GREEK SMALL LETTER IOTA WITH TONOS	ί
U+03B0	GREEK SMALL LETTER UPSILON WITH DIALYTIKA AND TONOS	ϋ
U+03B1	GREEK SMALL LETTER ALPHA	α
U+03B2	GREEK SMALL LETTER BETA	β
U+03B3	GREEK SMALL LETTER GAMMA	γ
U+03B4	GREEK SMALL LETTER DELTA	δ
U+03B5	GREEK SMALL LETTER EPSILON	ε
U+03B6	GREEK SMALL LETTER ZETA	ζ
U+03B7	GREEK SMALL LETTER ETA	η
U+03B8	GREEK SMALL LETTER THETA	θ
U+03B9	GREEK SMALL LETTER IOTA	ι
U+03BA	GREEK SMALL LETTER KAPPA	κ
U+03BB	GREEK SMALL LETTER LAMDA	λ
U+03BC	GREEK SMALL LETTER MU	μ
U+03BD	GREEK SMALL LETTER NU	ν
U+03BE	GREEK SMALL LETTER XI	ξ
U+03BF	GREEK SMALL LETTER OMICRON	ο
U+03C0	GREEK SMALL LETTER PI	π
U+03C1	GREEK SMALL LETTER RHO	ρ
U+03C2	GREEK SMALL LETTER FINAL SIGMA	ς
U+03C3	GREEK SMALL LETTER SIGMA	σ
U+03C4	GREEK SMALL LETTER TAU	τ
U+03C5	GREEK SMALL LETTER UPSILON	υ
U+03C6	GREEK SMALL LETTER PHI	φ
U+03C7	GREEK SMALL LETTER CHI	χ
U+03C8	GREEK SMALL LETTER PSI	ψ
U+03C9	GREEK SMALL LETTER OMEGA	ω

Codepoints	Name	Glyphe
U+03CA	GREEK SMALL LETTER IOTA WITH DIALYTIKA	ϊ
U+03CB	GREEK SMALL LETTER UPSILON WITH DIALYTIKA	ϋ
U+03CC	GREEK SMALL LETTER OMICRON WITH TONOS	ό
U+03CD	GREEK SMALL LETTER UPSILON WITH TONOS	ύ
U+03CE	GREEK SMALL LETTER OMEGA WITH TONOS	ώ

Für die Transliteration der griechischen Schriftzeichen in lateinische Schriftzeichen muss entsprechend des Übereinkommens über die Angabe von Familiennamen und Vornamen in den Personenstandsbüchern die einschlägige Norm [ISO 843] verwendet werden. Die Ergebnisse der Transliteration werden von konformen Implementationen dieser DIN SPEC vollständig unterstützt.

A.3. Kyrillische Buchstaben

Jedes IT-Verfahren sollte die in Tabelle A.2 aufgeführten Schriftzeichen unterstützen, die insbesondere für den Datenaustausch bulgarischer Namen (im weiteren Sinne) im europäischen Kontext benötigt werden. Die kyrillischen Schriftzeichen sollten unterstützt werden, um für den grenzüberschreitenden Datenaustausch bulgarische Namen im unveränderten Original übertragen zu können.

Tabelle A.2. Kyrillische Buchstaben

Codepoints	Name	Glyphe
U+040D	CYRILLIC CAPITAL LETTER I WITH GRAVE	Ѐ
U+0410	CYRILLIC CAPITAL LETTER A	А
U+0411	CYRILLIC CAPITAL LETTER BE	Б
U+0412	CYRILLIC CAPITAL LETTER VE	В
U+0413	CYRILLIC CAPITAL LETTER GHE	Г
U+0414	CYRILLIC CAPITAL LETTER DE	Д
U+0415	CYRILLIC CAPITAL LETTER IE	Е
U+0416	CYRILLIC CAPITAL LETTER ZHE	Ж
U+0417	CYRILLIC CAPITAL LETTER ZE	З
U+0418	CYRILLIC CAPITAL LETTER I	И
U+0419	CYRILLIC CAPITAL LETTER SHORT I	Й
U+041A	CYRILLIC CAPITAL LETTER KA	К
U+041B	CYRILLIC CAPITAL LETTER EL	Л
U+041C	CYRILLIC CAPITAL LETTER EM	М
U+041D	CYRILLIC CAPITAL LETTER EN	Н
U+041E	CYRILLIC CAPITAL LETTER O	О
U+041F	CYRILLIC CAPITAL LETTER PE	П
U+0420	CYRILLIC CAPITAL LETTER ER	Р
U+0421	CYRILLIC CAPITAL LETTER ES	С
U+0422	CYRILLIC CAPITAL LETTER TE	Т
U+0423	CYRILLIC CAPITAL LETTER U	У
U+0424	CYRILLIC CAPITAL LETTER EF	Ф

Codepoints	Name	Glyphe
U+0425	CYRILLIC CAPITAL LETTER HA	Х
U+0426	CYRILLIC CAPITAL LETTER TSE	Ц
U+0427	CYRILLIC CAPITAL LETTER CHE	Ч
U+0428	CYRILLIC CAPITAL LETTER SHA	Ш
U+0429	CYRILLIC CAPITAL LETTER SHCHA	Щ
U+042A	CYRILLIC CAPITAL LETTER HARD SIGN	Ъ
U+042C	CYRILLIC CAPITAL LETTER SOFT SIGN	Ь
U+042E	CYRILLIC CAPITAL LETTER YU	Ю
U+042F	CYRILLIC CAPITAL LETTER YA	Я
U+0430	CYRILLIC SMALL LETTER A	а
U+0431	CYRILLIC SMALL LETTER BE	б
U+0432	CYRILLIC SMALL LETTER VE	в
U+0433	CYRILLIC SMALL LETTER GHE	г
U+0434	CYRILLIC SMALL LETTER DE	д
U+0435	CYRILLIC SMALL LETTER IE	е
U+0436	CYRILLIC SMALL LETTER ZHE	ж
U+0437	CYRILLIC SMALL LETTER ZE	з
U+0438	CYRILLIC SMALL LETTER I	и
U+0439	CYRILLIC SMALL LETTER SHORT I	й
U+043A	CYRILLIC SMALL LETTER KA	к
U+043B	CYRILLIC SMALL LETTER EL	л
U+043C	CYRILLIC SMALL LETTER EM	м
U+043D	CYRILLIC SMALL LETTER EN	н
U+043E	CYRILLIC SMALL LETTER O	о
U+043F	CYRILLIC SMALL LETTER PE	п
U+0440	CYRILLIC SMALL LETTER ER	р
U+0441	CYRILLIC SMALL LETTER ES	с
U+0442	CYRILLIC SMALL LETTER TE	т
U+0443	CYRILLIC SMALL LETTER U	у
U+0444	CYRILLIC SMALL LETTER EF	ф
U+0445	CYRILLIC SMALL LETTER HA	х
U+0446	CYRILLIC SMALL LETTER TSE	ц
U+0447	CYRILLIC SMALL LETTER CHE	ч
U+0448	CYRILLIC SMALL LETTER SHA	ш
U+0449	CYRILLIC SMALL LETTER SHCHA	щ
U+044A	CYRILLIC SMALL LETTER HARD SIGN	ъ
U+044C	CYRILLIC SMALL LETTER SOFT SIGN	ь
U+044E	CYRILLIC SMALL LETTER YU	ю
U+044F	CYRILLIC SMALL LETTER YA	я

Codepoints	Name	Glyphe
U+045D	CYRILLIC SMALL LETTER I WITH GRAVE	ѝ

Für die Transliteration der kyrillischen Schriftzeichen in lateinische Schriftzeichen muss entsprechend des Übereinkommens über die Angabe von Familiennamen und Vornamen in den Personenstandsbüchern die einschlägige Norm [ISO 9] verwendet werden. Die Ergebnisse der Transliteration werden von konformen Implementationen dieser DIN SPEC vollständig unterstützt.

A.4. Nicht-Buchstaben E1

Jedes IT-Verfahren sollte die in Tabelle A.3 aufgeführten Schriftzeichen unterstützen, die zwar nicht für alle Namen, aber insbesondere für Namen juristischer Personen und für Produktnamen benötigt werden.

Tabelle A.3. Sonstige erweiterte Nicht-Buchstaben, erforderlich für nicht alle Namen

Codepoints	Name	Glyphe
U+0192	LATIN SMALL LETTER F WITH HOOK	f
U+02B0	MODIFIER LETTER SMALL H	h
U+02B3	MODIFIER LETTER SMALL R	r
U+02C6	MODIFIER LETTER CIRCUMFLEX ACCENT	^
U+02DC	SMALL TILDE	~
U+02E2	MODIFIER LETTER SMALL S	s
U+1D48	MODIFIER LETTER SMALL D	d
U+1D57	MODIFIER LETTER SMALL T	t
U+2018	LEFT SINGLE QUOTATION MARK	‘
U+201A	SINGLE LOW-9 QUOTATION MARK	‚
U+201C	LEFT DOUBLE QUOTATION MARK	“
U+201D	RIGHT DOUBLE QUOTATION MARK	”
U+201E	DOUBLE LOW-9 QUOTATION MARK	„
U+2020	DAGGER	†
U+2026	HORIZONTAL ELLIPSIS	...
U+2030	PER MILLE SIGN	‰
U+2039	SINGLE LEFT-POINTING ANGLE QUOTATION MARK	‹
U+203A	SINGLE RIGHT-POINTING ANGLE QUOTATION MARK	›
U+2070	SUPERSCRIFT ZERO	⁰
U+2074	SUPERSCRIFT FOUR	⁴
U+2075	SUPERSCRIFT FIVE	⁵
U+2076	SUPERSCRIFT SIX	⁶
U+2077	SUPERSCRIFT SEVEN	⁷
U+2078	SUPERSCRIFT EIGHT	⁸
U+2079	SUPERSCRIFT NINE	⁹
U+207F	SUPERSCRIFT LATIN SMALL LETTER N	ⁿ
U+2080	SUBSCRIPT ZERO	₀

Codepoints	Name	Glyphe
U+2081	SUBSCRIPT ONE	1
U+2082	SUBSCRIPT TWO	2
U+2083	SUBSCRIPT THREE	3
U+2084	SUBSCRIPT FOUR	4
U+2085	SUBSCRIPT FIVE	5
U+2086	SUBSCRIPT SIX	6
U+2087	SUBSCRIPT SEVEN	7
U+2088	SUBSCRIPT EIGHT	8
U+2089	SUBSCRIPT NINE	9
U+2122	TRADE MARK SIGN	™
U+221E	INFINITY	∞
U+2264	LESS-THAN OR EQUAL TO	≤
U+2265	GREATER-THAN OR EQUAL TO	≥

Anhang B. Technische Datentypen

B.1. Allgemeines

Im Rahmen von Schnittstellenvereinbarungen werden üblicherweise die Schriftzeichen, die ausgetauscht werden dürfen, abschließend bestimmt. Beispielsweise kann festgelegt werden, dass **ausschließlich** die normativen Schriftzeichen dieser DIN SPEC übermittelt werden dürfen. In Abschnitt E dieses Dokuments werden Formulierungen vorgeschlagen, die bei Schnittstellenvereinbarungen zum Einsatz kommen können. Entsprechende Einschränkungen können aus verschiedenen Gründen erfolgen:

- Aus fachlichen Gründen, um beispielsweise festzulegen, dass für Namen natürlicher Personen ausschließlich die in den Schriftzeichengruppen „Lateinische Buchstaben“ und „Nicht-Buchstaben N1“ enthaltenen Schriftzeichen zulässig sein sollen;

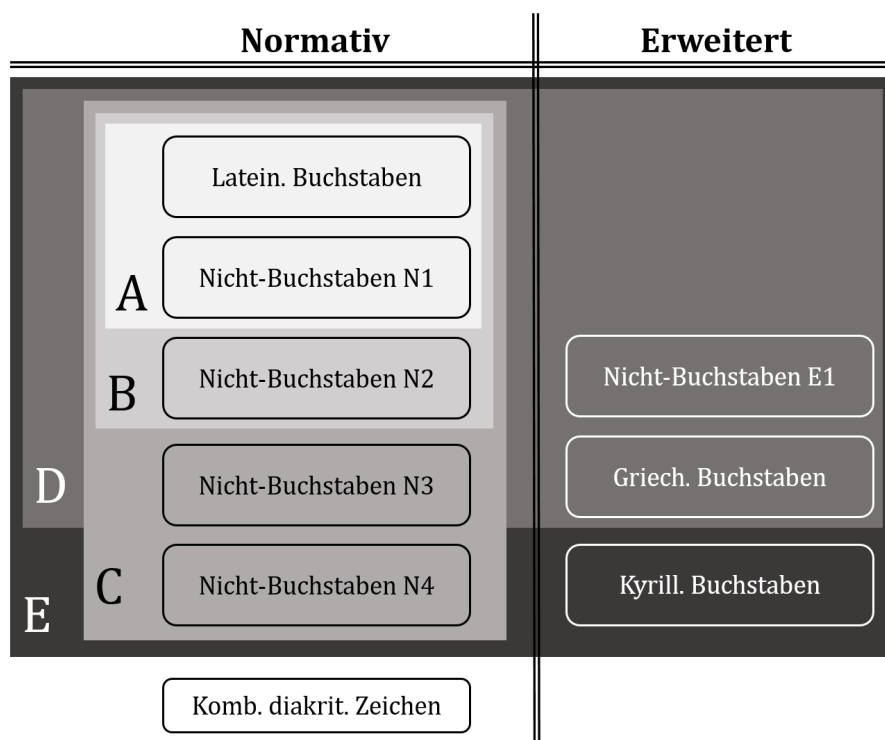
Für diesen Fall kann festgelegt werden, dass das zu übermittelnde Element vom Datentyp A (Abschnitt B.2) ist;

- Aus Gründen der Sicherheit, um beispielsweise durch Einschränkung auf bestimmte Schriftzeichengruppen auszuschließen, dass optisch identische, informationstechnisch aber unterschiedliche Zeichen verarbeitet werden. Auch die Verarbeitung unsichtbarer Zeichen wie beispielsweise dem breitenlosen Leerzeichen (ZERO WIDTH SPACE, U+200B) oder dem weichen Trennzeichen (SOFT HYPHEN, U+00AD) kann ausgeschlossen werden, weil vergleichbare Sicherheitsrisiken bestehen.

Für diesen Fall kann festgelegt werden, dass das zu übermittelnde Element vom Datentyp C (Abschnitt B.4) oder E (Abschnitt B.6) ist.

Um diese Ziele zu unterstützen, wird empfohlen, folgende Datentypen technisch umzusetzen (z. B. in Form von XSD Schema Datentypen). Dabei ist zu berücksichtigen, dass die Datentypen D (Abschnitt B.5) und E (Abschnitt B.6) nur von solchen IT-Verfahren implementiert werden können, die zusätzlich zu den normativen Schriftzeichen auch die empfohlenen Schriftzeichen des Anhangs A unterstützen. Die verbindliche Vorgabe eines der beiden Datentypen D oder E, beispielsweise in Schnittstellenvereinbarungen, kann daher nur dann empfohlen werden, wenn sichergestellt ist, dass alle am Datenaustausch beteiligten IT Verfahren nicht nur konform zur DIN SPEC 91379 sind, sondern darüber hinaus auch die entsprechenden Schriftzeichengruppen des Anhangs A unterstützen.

Informativ ist dieser DIN SPEC eine Implementierung der nachfolgend beschriebenen Datentypen in der Datei `din-91379-datatypes.xsd` beigefügt. Jeder der in den Abschnitten B.2 bis B.6 beschriebenen Datentypen ist darin jeweils als `xs:simpleType` realisiert. Es handelt sich jeweils um eine Einschränkung des Basistyps `xs:string` durch Festlegung eines regulären Ausdrucks, der ausschließlich die für den jeweiligen Datentyp vorgesehenen Schriftzeichen zulässt. Die regulären Ausdrücke setzen zudem die Anforderung um, dass kombinierende diakritische Zeichen ausschließlich in den Zeichensequenzen genutzt werden, die in Tabelle 3 abschließend genannt sind. Die informativ beigefügte Implementierung in XML Schema geht davon aus, dass die zu validierenden Daten NFC normalisiert sind (siehe Abschnitt B.7).

Abbildung B.1. Datentypen der DIN SPEC 91379**Legende**

A	Datentyp A: für Namen natürlicher Personen
B	Datentyp B: für sonstige Namen, wie z.B. Straßennamen mit Hausnummer
C	Datentyp C: für alle normative Schriftzeichen
D	Datentyp D: für Namen juristischer Personen
E	Datentyp E: für alle normativen und erweiterten Schriftzeichen

Bild B.1 und Tabelle B.1 zeigen, wie die in der DIN SPEC zuvor definierten Schriftzeichengruppen zu technischen Datentypen zusammengefügt werden. Die mit den Buchstaben A bis E benannten technischen Datentypen enthalten die Schriftzeichen aller über ihrer Fläche liegenden Schriftzeichengruppen. Beispielsweise enthält der Datentyp D die Schriftzeichengruppen „Lateinische Buchstaben“, „Nicht-Buchstaben N1“, „Nicht-Buchstaben N2“, „Nicht-Buchstaben N3“, „Nicht-Buchstaben E1“ sowie „Griechische Buchstaben“.

Tabelle B.1. Übersicht über die empfohlenen Datentypen und zugehörigen Schriftzeichengruppen

Schriftzeichengruppe	Tabelle	Datentyp A	Datentyp B	Datentyp C	Datentyp D	Datentyp E
Latein. Buchstaben (normativ)	3	X	X	X	X	X
Nicht-Buchstaben N1 (normativ)	5	X	X	X	X	X
Nicht-Buchstaben N2 (normativ)	6		X	X	X	X
Nicht-Buchstaben N3 (normativ)	7			X	X	X

Schriftzeichengruppe	Tabelle	Datentyp A	Datentyp B	Datentyp C	Datentyp D	Datentyp E
Nicht-Buchstaben N4 (normativ)	8			X		X
Komb. diakrit. Zeichen (normativ)	4					
Griech. Buchstaben (erweitert)	A.1				X	X
Kyrill. Buchstaben (erweitert)	A.2					X
Nicht-Buchstaben E1 (erweitert)	A.3				X	X

Die kombinierenden diakritischen Zeichen der Tabelle 4, „Normative kombinierende diakritische Zeichen“ kommen nicht alleinstehend in den Datentypen vor, weil dann beliebige Kombinationen mit diesen Zeichen gebildet werden könnten. Diese sind jedoch alle ein Bestandteil abschließend festgelegter Kombinationen in Zeichensequenzen.

B.2. Datentyp A

Der Datentyp A gibt wieder, welche Schriftzeichen in hoheitlichen Dokumenten für Namen natürlicher Personen verwendet werden.

Datentyp A ist definiert durch die abschließenden Listen aller lateinischen Buchstaben nach Tabelle 3 und der für alle Namen erforderlichen Nicht-Buchstaben nach Tabelle 5. Andere Schriftzeichen werden durch diesen Datentyp zurückgewiesen.

B.3. Datentyp B

Der Datentyp B wurde vor allem für sonstige Namen, wie z. B. Ortsnamen und Straßennamen mit Hausnummer, entworfen.

Datentyp B ist definiert durch die abschließenden Listen aller lateinischen Buchstaben nach Tabelle 3, der für alle Namen erforderlichen Nicht-Buchstaben nach Tabelle 5 und der sonstigen normativen Nicht-Buchstaben nach Tabelle 6. Andere Schriftzeichen werden durch diesen Datentyp zurückgewiesen.

B.4. Datentyp C

Der Datentyp C wurde für alle normativen Schriftzeichen der DIN SPEC entworfen. Er ist somit die technische Umsetzung der Schnittstellenvereinbarung „Alle nach DIN SPEC 91379 normativen Schriftzeichen“. Texte mit griechischen oder kyrillischen Buchstaben oder mit erweiterten (nicht-normativen) Nicht-Buchstaben sind unzulässig.

Datentyp C ist definiert durch die abschließenden Listen aller lateinischen Buchstaben nach Tabelle 3, der für alle Namen erforderlichen Nicht-Buchstaben nach Tabelle 5, der sonstigen normativen Nicht-Buchstaben nach Tabelle 6, der normativen Nicht-Buchstaben, erforderlich für Namen juristischer Personen nach Tabelle 7 und der normativen für Namen ungeeigneten Nicht-Buchstaben nach Tabelle 8. Andere Schriftzeichen werden durch diesen Datentyp zurückgewiesen.

B.5. Datentyp D

Dieser Datentyp kann nur von solchen IT Verfahren unterstützt werden, die auch die nicht-normativen Schriftzeichen in Anhang A der DIN SPEC 91379 unterstützen. Er wurde vor allem für Namen juristischer Personen und für Produkt-namen entworfen.

Datentyp D ist definiert durch die abschließenden Listen aller lateinischen und griechischen Buchstaben nach den Tabellen 3 und A.1, der für alle Namen erforderlichen Nicht-Buchstaben nach Tabelle 5 sowie der für Namen (im weiteren Sinne) geeigneten normativen und erweiterten Nicht-Buchstaben nach den Tabellen 6, 7 und A.3. Andere Schriftzeichen werden durch diesen Datentyp zurückgewiesen.

B.6. Datentyp E

Dieser Datentyp kann nur von solchen IT Verfahren unterstützt werden, die auch die nicht-normativen Schriftzeichen in Anhang A der DIN SPEC 91379 unterstützen. Er wurde für alle normativen und erweiterten Schriftzeichen der DIN SPEC entworfen. Ein Einsatzgebiet dieses Datentyps kann der grenzüberschreitende Datenaustausch sein, wenn auch griechische und kyrillische Buchstaben benötigt werden. Er ist somit die technische Umsetzung der Schnittstellenvereinbarung „Alle nach DIN SPEC 91379 normativen und nicht-normativen Schriftzeichen“. Texte mit Buchstaben oder Nicht-Buchstaben, die in der DIN SPEC nicht enthalten sind, wie z. B. asiatische oder arabische Buchstaben, sind unzulässig.

Datentyp E ist definiert durch die abschließenden Listen aller lateinischen, griechischen und kyrillischen Buchstaben nach den Tabellen 3, A.1 und A.2 sowie aller normativen und erweiterten Nicht-Buchstaben nach den Tabellen 5, 6, 7, 8 und A.3. Andere Schriftzeichen werden durch diesen Datentyp zurückgewiesen.

B.7. Unicode Normalisierung NFC

Viele abstrakte Zeichen des Unicode Standards können auf unterschiedliche Arten repräsentiert werden. So kann zum Beispiel das abstrakte Zeichen A-Umlaut sowohl als Sequenz mit einem einzigen Schriftzeichen U+00C4 *LATIN CAPITAL LETTER A WITH DIAERESIS*, als auch als Zeichensequenz <U+0041, U+0308> der beiden Schriftzeichen *LATIN CAPITAL LETTER A* und *COMBINING DIAERESIS* repräsentiert werden. Beide Repräsentationen (*Encodings*) sind *äquivalent* im Sinne des Abschnitts 2.12 “Equivalent Sequences” des Unicode Standards.

Dieser Sachverhalt kann den Umgang mit Zeichenketten erschweren, die auf dem Unicode Standard basieren. Insbesondere können Vergleiche von Zeichenketten kompliziert werden, wenn an Stelle der Gleichheit die Äquivalenz geprüft werden muss. Eine ausführliche Darstellung sprengt den Rahmen dieses Dokuments, man findet sie beispielsweise in [W3C CHARMOD-NORM].

Daher kann es insbesondere für den Datenaustausch zwischen IT Verfahren sachgerecht sein, in Schnittstellenvereinbarungen *eindeutige* Repräsentationen von Zeichensequenzen festzulegen. In Abschnitt 3.11 des Unicode Standards werden zu diesem Zweck vier Normalformen definiert (Definitionen D118 bis D121). Die meistverbreitete Normalform in Anwendungen ist NFC. Es wird empfohlen, die Anwendung der Normalform NFC in Schnittstellenbeschreibungen vorzugeben, um eine eindeutige Repräsentation von Unicode Zeichensequenzen zu gewährleisten.

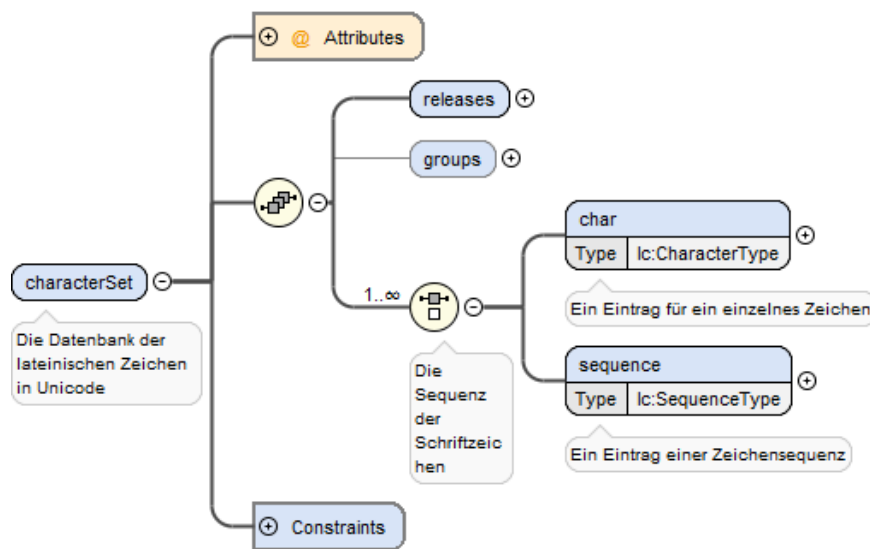
Die Empfehlung zu Gunsten der Normalform NFC wird auch durch die Eignung für Legacy Zeichensätze begründet. Sofern Daten gemäß ISO-8859 oder DIN 66303 gespeichert sind, liegen sie bereits in NFC Normalform vor, denn diese Legacy Zeichensätze sind *pränormalisiert* im Sinne der Definition UAX15-D2 [<https://unicode.org/reports/tr15/#UAX15-D2>] von [Unicode 15].

Die informell zur Verfügung gestellte Implementierung der Datentypen A bis E als XML Schema wurde entsprechend dieser Empfehlung realisiert. Der für jeden Datentyp erstellte reguläre Ausdruck setzt jeweils voraus, dass die zu validierende Zeichenkette in der NFC Normalform vorliegt. Für das oben genannte Beispiel bedeutet das: eine Zeichenkette, in der das abstrakte Zeichen A Umlaut enthalten ist, muss dieses Zeichen als Codepoint U+004C enthalten. Die ebenfalls mögliche Repräsentation <U+0041, U+0308> ist zwar äquivalent im Sinne des Unicode Standard, aber gleichwohl nicht valide bezüglich der informell bereitgestellten Umsetzung der Datentypen.

Anhang C. Strukturierte und tabellarische Übersicht aller Zeichen

Informativ ist der DIN SPEC eine Datei `latinchars.xml` mit allen enthaltenen Schriftzeichen im Format XML 1.0 beigefügt. Sie ist valide bezüglich der ebenfalls beigefügten Datei `latinchars.xsd` im Format XML Schema 1.0. Die grundsätzliche Struktur ist in Bild C.1 dargestellt. Detaillierte Informationen sind der Dokumentation des XSD Schema in `latinchars.xsd` zu entnehmen.

Abbildung C.1. Struktur der Datenbasis aller Schriftzeichen dieser DIN SPEC



Nach erläuternden Angaben zu `releases` dieser Datei folgt die Beschreibung definierter Schriftzeichengruppen im Element `groups`. Anschließend folgt eine Sequenz von Schriftzeichen, welches jeweils entweder ein kodiertes Zeichen im Element `char` oder eine kodierte Zeichensequenz im Element `sequence` ist.

Jedes Element `char` oder `sequence` verfügt über ein Kindelement `cp` mit dem Codepoint bzw. den Codepoints des jeweiligen Schriftzeichens. Jeder Codepoint ist mit genau vier Hexadezimalziffern angegeben. Im Falle kodierter Zeichensequenzen sind die zugehörigen Codepoints durch genau ein Leerzeichen `U+0020` getrennt. Der Name des Schriftzeichens ist jeweils im Kindelement `name` angegeben.

Beispielhaft ist je ein Element `char` und `sequence` für die Schriftzeichen `A` (`U+0041`) und `Á` (`<U+0041, U+030B>`) mit den beiden wesentlichen Kindelementen `cp` und `name` dargestellt. Beide Schriftzeichen gehören der als `b11` bezeichneten Schriftzeichengruppe Normative lateinische Buchstaben (siehe Tabelle 3) an:

```
<char group="b11">
  <cp>0041</cp>
  <name>LATIN CAPITAL LETTER A</name>
</char>
<sequence group="b11">
  <cp>0041 030B</cp>
  <name>LATIN CAPITAL LETTER A WITH COMBINING DOUBLE ACUTE ACCENT</name>
</sequence>
```

Die Datei enthält nur wenige weitere Angaben zum jeweiligen Schriftzeichen. Alle Eigenschaften des jeweiligen Schriftzeichens sollen der vom Unicode Konsortium bereitgestellten Unicode Character Database (UCD) entnommen werden, die in [Unicode 42] beschrieben ist.

Informativ ist der DIN SPEC zudem eine tabellarische Übersicht aller enthaltenen Schriftzeichen beigefügt, siehe 2018-11-15_din-spec-91379.xlsx.

Anhang D. Abbildung auf Legacy Codepages

Informativ ist der DIN SPEC eine tabellarische Übersicht von Abbildungen der Schriftzeichen dieser DIN SPEC auf die Zeichen der Legacy Codepages ISO 8859-1, ISO 8859-15, CP1252 und EBCDIC CP1141 beigefügt, siehe 2018-11-15_legacy-mappings.xlsx.

Anhang E. Häufig gestellte Fragen

E.1. Allgemeine Fragen zur DIN SPEC 91379

E.1.1. Welches Problem soll durch die DIN SPEC 91379 gelöst werden?

Die Identifikation von Personen in IT-Verfahren der öffentlichen Verwaltung erfolgt regelhaft auf der Grundlage ihres Namens und weiterer persönlicher Daten. Schon aus diesem Grund muss sichergestellt werden, dass Namen von Personen in allen IT-Verfahren auf die gleiche Weise verarbeitet werden können. Darüber hinaus haben Menschen einen Rechtsanspruch auf die korrekte Schreibweise ihres Namens.

Die DIN SPEC 91379 legt einen Mindestzeichensatz für diesen Zweck fest, der von allen IT-Verfahren der öffentlichen Verwaltung unterstützt werden muss.

E.1.2. Was wird durch die Spezifikation festgelegt?

Die Spezifikation definiert im Wesentlichen eine Menge von Zeichen, die jedes IT-Verfahren, für das die Entscheidung 2019/16 des IT-Planungsrats einschlägig ist, vollständig unterstützen muss. Es handelt sich um Buchstaben, Ziffern, Interpunktions- und weitere Zeichen.

E.1.3. Was ist Unicode?

Unicode ist ein universeller Standard, der Zeichen auf Zahlen abbildet (kodiert) und sie somit für einen Computer erst „lesbar“ macht. Er ist außerdem ein übergeordnetes Regelwerk zum Umgang mit den kodierten Zeichen.

Es gibt auf der Welt viele verschiedene Schriftsysteme (z. B. Lateinisch, Griechisch, Kyrillisch). Für die meisten davon gibt es wiederum – historisch gewachsen – viele unterschiedliche Zuordnungen der im Schriftsystem enthaltenen Zeichen zu Zahlen (die so genannten Zeichenkodierungen, z. B. ISO 8859-2, CP1250). Aus dem Nebeneinander vieler hundert verschiedener Kodierungssysteme für diverse Schriftsysteme bzw. Alphabete resultieren jedoch Interoperabilitätsprobleme auf unterschiedlichen Ebenen, da dieselben Zahlen für unterschiedliche Zeichen stehen können.

Im Zuge der Globalisierung hat sich die IT-Industrie dieser Problematik angenommen und eine Lösung entwickelt. Der Unicode-Standard umfasst alle gebräuchlichen Alphabete und weist jedem Zeichen eine eindeutige Zahl (seinen Codepoint) zu. Der Unicode-Kernstandard (bzw. die inhaltlich identische ISO-Norm 10646, auch *Universal Coded Character Set* [UCS]) ist der international anerkannte Standard für die Zeichenkodierung und geeignet, die oben genannten Interoperabilitätsprobleme zu lösen. Unicode wird von allen modernen Betriebssystemen und systemnahen Anwendungen, wie z. B. Datenbanken, unterstützt.

Nähere Informationen über den Standard Unicode erhalten Sie auf der Webseite www.Unicode.org [<http://www.Unicode.org>] des Unicode-Konsortiums.

E.1.4. Im Zusammenhang mit Zeichensätzen wird immer wieder von UTF-8 (oder UTF-16) gesprochen. Ist das das gleiche wie Unicode?

Nein. UTF bedeutet *UCS Transformation Format*. Es ist eine Methode, die Codepoints der Unicode-Zeichen auf Folgen von Bytes abzubilden. Die Abbildung auf Bytes (Encoding) wird benötigt, um in IT-Systemen Unicode-Zeichen verarbeiten, übertragen und speichern zu können.

Der Unicode-Standard trifft auch Aussagen über verschiedene UTF-Encoding. In jedem der Formate lassen sich alle im Unicode-Standard enthaltenen Zeichen (bzw. deren Codepoints) darstellen. Auch lässt sich jedes dieser Formate verlustfrei in die jeweils anderen UTF-Varianten konvertieren.

Die Formate unterscheiden sich vor allem hinsichtlich ihres Speicherplatzbedarfs. Bei Texten, welche hauptsächlich aus lateinischen Buchstaben bestehen, ist UTF-8 in der Regel ressourcenschonender im Speicherbedarf als UTF-16. Die Wahl des geeigneten Encodings, besonders zur internen Zeichenkettenrepräsentation, ist ein Detail der technischen Umsetzung, welches bei Schnittstellenvereinbarungen relevant sein kann. Die DIN SPEC 91379 trifft diesbezüglich keine Aussagen.

- E.1.5.** Weshalb muss eine eigene Spezifikation entwickelt werden? Hätte nicht einfach die vollständige Umsetzung des Unicode-Standards gefordert werden können?

Im Abschnitt „Abwägung von Alternativen“ auf der Seite 6 wird begründet, warum eine abschließende Aufzählung der zu unterstützenden Zeichen notwendig ist anstelle der Forderung, alle (lateinischen) Unicode-Zeichen zu unterstützen.

- E.1.6.** Wie ist die Beziehung zwischen der DIN SPEC 91379 und dem internationalen Standard Unicode?

Die DIN SPEC 91379 basiert vollständig auf Unicode. Sie zeichnet lediglich eine Teilmenge der durch den Unicode-Standard festgelegten Schriftzeichen und Zeichensequenzen aus.

- E.1.7.** Was passiert bei einer Fortschreibung von Unicode? Führt das automatisch zu einer neuen Version der Spezifikation?

Nein. Es stimmt, dass der Standard Unicode immer weiter entwickelt wird, aber die DIN SPEC 91379 basiert auf einem Kernbereich von Unicode, der stabil ist. Es ist nicht damit zu rechnen, dass es bei einer neuen Version von Unicode Veränderungen in diesem Kernbereich geben wird, so dass es sehr wahrscheinlich nicht notwendig sein wird, deshalb eine neue Fassung der Spezifikation herauszugeben.

Unabhängig vom Standard Unicode kann es aber zu einer Weiterentwicklung der Spezifikation kommen.

- E.1.8.** Weshalb macht Deutschland eigene Vorgaben? Wäre es nicht sinnvoll, sich auf europäischer Ebene auf eine Norm zu einigen?

Ja, das wäre sinnvoll. Deutschland hat einen entsprechenden Normierungsantrag an die europäischen Normierungsgremien gerichtet. Dieser ist jedoch von der Europäischen Kommission zunächst leider abgelehnt worden. Deshalb wird nun der Weg von der DIN SPEC über eine DIN-Norm hin zu einer europäischen CEN-Norm beschritten.

- E.1.9.** Ist die DIN SPEC 91379 auch in englischer Sprache verfügbar?

Derzeit leider nicht. Es besteht jedoch die Absicht, Anfang 2020 eine englischsprachige Version der DIN SPEC 91379 zu veröffentlichen.

- E.1.10.** Warum wurde eine DIN SPEC an Stelle einer DIN-Norm erstellt?

Es gab einen hohen Bedarf an einer schnellen Verfügbarkeit eines Nachfolgers von String.Latin 1.1. Eine DIN SPEC kann schneller erstellt werden als eine DIN-Norm.

- E.1.11.** Wie ist der weitere Prozess der Normung geplant?

Im nächsten Schritt soll eine DIN-Norm erstellt werden. Unterhalb des DIN-Normenausschusses NA 043-03-03 AA „Elektronisches Geschäftswesen“ wurde ein neuer DIN-Arbeitskreis (DIN-AK) „Elektronische Verarbeitung von Namen“ für die Erstellung der Norm gegründet. Im August 2019 hat der DIN-AK seine Arbeit aufgenommen.

Da es bereits bei der Erstellung der DIN SPEC 91379 ein umfangreiches Beteiligungsverfahren gab (eine Vor-Version ist vom DIN zwecks Kommentierung durch die Öffentlichkeit im Herbst 2018 veröffentlicht worden), darf angenommen werden, dass die DIN-Norm inhaltlich der DIN SPEC 91379 weitgehend entsprechen wird. Es wird eine Veröffentlichung der DIN-Norm Ende 2020 angestrebt.

Direkt daran anschließend soll die Erstellung einer CEN-Norm begonnen werden.

Wenn Sie Interesse haben, sich an der Normierung zu beteiligen, können Sie entweder noch nachträglich dem DIN-AK beitreten oder an der öffentlichen Kommentierung des Normentwurfes teilnehmen. In beiden Fällen senden Sie bitte eine formlose E-Mail an <kosit@finanzen.bremen.de>.

E.1.12. Wie unterscheidet sich die DIN SPEC 91379 von ihrem Vorgänger [String Latin 1.1]?

1. **Vollständigkeit:** Die DIN SPEC 91379 berücksichtigt alle Transliterationsempfehlungen der ISO zur Transliteration aus anderen Schriftsystemen in die lateinische Schrift. Sie berücksichtigt zudem die Rückmeldungen der deutschen Behörden mit dem Vorgängerstandard seit 2012. Dazu gehören insbesondere die Erfahrungen aus der Nacherfassung von Personenstandsurkunden.

In der DIN SPEC 91379 sind insgesamt 753 Schriftzeichen normativ vorgegeben. Das sind 263 Schriftzeichen mehr als in String.Latin 1.1 mit insgesamt 490 Zeichen.

2. **Klare Konformitätsregeln:** Die Regelungen zur Konformität von IT-Verfahren zu String.Latin 1.1 wurden häufig als zu unbestimmt kritisiert. Die DIN SPEC 91379 ist präziser, indem festgelegt wird, dass die normativen Schriftzeichen bei allen für Namen (im weiteren Sinne) bestimmten Datenfeldern verarbeitbar sein müssen (siehe Abschnitt 4.1).
3. **Systematischer Aufbau durch Schriftzeichengruppen:** Die Schriftzeichen der DIN SPEC sind sechs normativen und drei empfohlenen Schriftzeichengruppen gemäß Tabelle 1 zugeordnet. Dies dient zunächst der Transparenz und Verständlichkeit der Spezifikation. Darüber hinaus ermöglicht diese Systematik die Bildung der Datentypen in Anhang B der DIN SPEC 91379.

Der Vorgänger String.Latin 1.1 hat keinen vergleichbaren Mechanismus.

4. **Differenzierte Datentypen:** Die DIN SPEC 91379 empfiehlt insgesamt fünf Datentypen für unterschiedliche fachliche Anforderungen (siehe Abschnitt B). Diese entsprechen unterschiedlichen Kombinationen von Schriftzeichengruppen.

String.Latin 1.1 verfügt über keine Schriftzeichengruppen und referenziert daher auch nur auf einen technischen Datentyp.

5. **Abbildung auf Grundbuchstaben:** Pässe und Ausweise, wie beispielsweise der deutsche Personalausweis, verfügen in der Regel über eine maschinenlesbare Zone (MRZ), in der der Name des Inhabers gemäß der international abgestimmten Regeln von ICAO ausschließlich mit Großbuchstaben A...Z geschrieben werden darf. Das Regelwerk von ICAO deckt jedoch nicht alle Buchstaben ab, die in der DIN SPEC aufgeführt werden. Deshalb wird von der DIN SPEC im Abschnitt 5 eine normative Abbildung aller Buchstaben auf die Grundbuchstaben mitgeliefert. Ein weiterer relevanter Anwendungsfall einer solchen Abbildung ist die Verwendung von Suchformen. Für eine höhere Effizienz und um Namen auch dann identifizieren zu können, wenn sie ausschließlich mit dem Zeichenvorrat von US-ASCII geschrieben wurden, werden vorhandene Datensätze und Suchanfragen entsprechend des Regelwerks vereinfacht. Eine einheitliche Abbildungsvorschrift reduziert Fehler beim Identifizieren von Namen.

Für den Vorgänger [String Latin 1.1] gab es ebenfalls eine solche Abbildungsvorschrift, allerdings in dem separaten Dokument [AKI Identifikation].

Für die MRZ gilt, dass Leerzeichen, Komma und Bindestrich als „<“ umgesetzt werden. Für Suchformen gibt es unterschiedliche Strategien im Umgang mit Nicht-Buchstaben. Zum Umgang mit Nicht-Buchstaben macht die DIN SPEC keine Vorgaben.

6. **Europäischer Datenaustausch:** Die DIN SPEC 91379 nennt im erweiterten Bereich kyrillische und griechische Buchstaben, die von IT-Verfahren insbesondere im Kontext europäischer Verbundverfahren unterstützt werden *sollten*. Bei vollständiger Unterstützung auch des erweiterten Bereichs der DIN SPEC 91379 werden alle europäischen Amtssprachen abgedeckt.

String.Latin 1.1 hat keinen erweiterten Bereich mit anderen als lateinischen Schriftzeichen.

E.1.13. Was bedeutet NFC?

NFC steht für „Normalization Form C“ und ist eine Normalisierungsform für Unicode-Zeichen, da es in Unicode teilweise mehrere Möglichkeiten gibt, ein Zeichen darzustellen.

Technisch werden hierbei die Codepoints zuerst zu einer kanonisch äquivalenten Form dekomponiert und danach wieder so weit wie möglich zu existierenden Codepoints rekombiniert („Canonical Decomposition, followed by Canonical Composition“). Das Ergebnis einer Normalisierung in NFC behält immer die visuelle Darstellung und Semantik der Zeichenfolge bei, der Mensch erkennt hierbei also keinen Unterschied. Intern liegen Zeichenketten jedoch danach in einer einheitlichen Form vor und können so – bei konsistenter Benutzung der Normalisierung – sortiert und verglichen werden.

Für mehr Informationen siehe Abschnitt B.7 in diesem Dokument, sowie den offiziellen Technical Report [AKI Identifikation] oder die [FAQ Normalization] des Unicode Konsortiums.

E.1.14. Was ist ein Legacy-Zeichensatz bzw. eine Legacy Codepage?

Als Legacy Codepage werden gemeinhin Codepages bezeichnet, welche vor Unicode existiert haben und nicht alle Zeichen, welche im Unicode-Standard vorhanden sind, abdecken können. Dies ist meist durch die technische Einschränkung von genau einem Byte pro Zeichen (so genannte Single Byte Character Sets, SBCS) bedingt. Beispiele für Legacy Codepages sind ASCII, ISO 8859-1 (Latin 1), ISO 8859-15 (Latin 9), CP1252 (Windows-1252) oder IBM EBCDIC 1141.

Muss eine unicodefähige Anwendung Daten an eine nicht-unicodefähige Schnittstelle weitergeben, kann es passieren, dass durch das Nichtvorhandensein einzelner Zeichen in der Legacy Codepage Daten verloren gehen. Grundsätzlich kann hier entweder substituiert werden, also jegliche Zeichen außerhalb des Zeichenumfangs der Zielcodepage werden gegen ein Substitutionszeichen ersetzt, oder es wird versucht, nicht darstellbare Zeichen auf ähnliche, in der Zielcodepage vorhandene Zeichen zu mappen. Hierfür bietet die DIN SPEC einen Vorschlag des Mappings der Zeichen in String.Latin+ auf häufig verwendete SBCS, siehe Abschnitt D.

E.2. Verbindlichkeit für die öffentliche Verwaltung

E.2.1. Ist die DIN SPEC 91379 für IT-Verfahren der öffentlichen Verwaltung verbindlich?

Ja. Der IT-Planungsrat hat am 12. März 2019 mit der Entscheidung 2019/16 die DIN SPEC 91379 verbindlich für alle IT-Verfahren erklärt, die dem Bund-Länder übergreifenden Datenaustausch oder dem Datenaustausch mit Bürgern und Wirtschaft dienen. Das Anwendungsgebiet der Spezifikation ist die Verarbeitung von Namen (im weiteren Sinne, also Namen natürlicher und juristischer Personen sowie Namen von Objekten wie beispielsweise Orten, Straßen oder Produkten). Zum 1. Februar 2020 und 2022 müssen die Mitglieder des IT-Planungsrates über den Stand der Umsetzung der Entscheidung berichten.

Die Umsetzungsfrist wurde am 23. Oktober 2019 mit der Entscheidung 2019/53 des IT-Planungsrates auf den 1. November 2024 festgelegt.

Der Vorgänger der DIN SPEC 91379 (intern: String.Latin+ 1.2) ist der von der KoSIT herausgegebene Standard [String Latin 1.1]. Diese Version wurde vom IT-Planungsrat mit der Entscheidung 2014/04 mit einer Umsetzungsfrist bis zum Jahr 2017 für verbindlich erklärt.

Neben der Herleitung der Verbindlichkeit über Beschlüsse des IT-Planungsrates gibt es für String.Latin+ noch eine weitere spezielle Rechtsgrundlage: das „*Übereinkommen über die Angabe von Familiennamen und Vornamen in den Personenstandsbüchern*“ vom 13. September 1973 (BGBl. 1976 II, Seite 1473 f.). Demnach sind Namen in nicht lateinischer Schrift „ohne Übersetzung soweit wie möglich durch Transliteration wiederzugeben. Sind von der Internationalen Normenorganisation (ISO) empfohlene Normen vorhanden, so sind sie anzuwenden“. Sofern die Namen in lateinischen Schriftzeichen vorliegen, „so sind diese Familiennamen und Vornamen buchstabengetreu ohne Änderung oder Übersetzung wieder-

zugeben“. Diakritische Zeichen sind dementsprechend unverändert wiederzugeben, selbst wenn die deutsche Sprache diese Zeichen nicht kennt.

Da Deutschland diesem internationalen Übereinkommen beigetreten ist, sind dessen Festlegungen für das Personenstandswesen direkt bindend. Indirekt ist davon auch das Pass- und Ausweiswesen betroffen. Daraus ergibt sich wiederum eine Relevanz für alle staatlichen Leistungen, bei denen Namen in Pässen und Ausweisen unverändert wiedergegeben werden müssen, z. B. für Schulzeugnisse oder Führungszeugnisse.

Die DIN SPEC 91379 beschreibt in ihrem normativen Teil die Zeichenmenge, die unterstützt werden muss, um die Anforderungen des „*Übereinkommens über die Angabe von Familiennamen und Vornamen in den Personenstandsbüchern*“ zu erfüllen.

E.2.2. Woraus ergibt sich die Verbindlichkeit?

Die Verbindlichkeit ergibt sich aus den Entscheidungen 2014/04 und 2019/16 des IT-Planungsrats.

Die Verbindlichkeit der Entscheidungen des IT-Planungsrats wiederum ergibt sich aus einer Änderung des Grundgesetzes (Artikel 91c GG). Darin heißt es unter anderem „Bund und Länder können auf Grund von Vereinbarungen die für die Kommunikation zwischen ihren informationstechnischen Systemen notwendigen Standards und Sicherheitsanforderungen festlegen.“ Zur Umsetzung des Artikels 91c GG haben Bund und Länder den [IT-Staatsvertrag] geschlossen. Darin wurde die Errichtung des IT-Planungsrats festgelegt, der verbindliche fachunabhängige und fachübergreifende IT-Interoperabilitäts- und IT-Sicherheitsstandards für Bund und Länder beschließt. String.Latin ist ein solcher fachunabhängiger IT-Interoperabilitätsstandard.

E.2.3. Ist der Beschluss des IT-Planungsrats auch für Kommunen bindend?

Die Entscheidung 2019/16 ist eine Festlegung von IT-Interoperabilitäts- und IT-Sicherheitsstandards gemäß § 3 des IT-Staatsvertrags. Solche Beschlüsse des IT-Planungsrats entfalten Bindungswirkung und werden vom Bund und den Ländern innerhalb jeweils vom IT-Planungsrat festzusetzender Fristen in ihren jeweiligen Verwaltungsräumen umgesetzt. Durch diese Umsetzung innerhalb des jeweiligen Bundeslandes wird die Entscheidung 2019/16 auch für Kommunen bindend.

E.2.4. In der Entscheidung 2014/04 des IT-Planungsrats zu String.Latin 1.1 gab es eine Einschränkung der Verbindlichkeit: „wenn nicht zwingende fachliche oder wirtschaftliche Gründe dagegen sprechen“. Warum wird diese Ausnahmeregelung nicht auch in der Entscheidung 2019/16 zur DIN SPEC 91379 (String.Latin + 1.2) erwähnt?

Diese Passage der Entscheidung 2014/04 wurde sehr oft missverstanden. In der Ziffer 2 wurde festgelegt, dass der Standard für alle „IT-Verfahren, die dem bund-länderübergreifenden Datenaustausch oder dem Datenaustausch mit Bürgern und Wirtschaft dienen“, umgesetzt werden muss – mit zwei verschiedenen Fristen. In der Ziffer 3 heißt es anschließend: „Die Mitglieder des IT-Planungsrats tragen in ihrer jeweiligen Gebietskörperschaft dafür Sorge, dass, sobald möglich, sämtliche IT-Verfahren konform zu diesem Standard sind, wenn nicht zwingende fachliche oder wirtschaftliche Gründe dagegen sprechen.“ Weitergehend zur Ziffer 2 sind hier alle IT-Verfahren gemeint, also auch jene, die nicht dem bund-länderübergreifenden Datenaustausch oder dem Datenaustausch mit Bürgern und Wirtschaft dienen. Die konforme Umstellung dieser IT-Verfahren steht unter dem Vorbehalt fachlicher oder wirtschaftlicher Einwände.

Die Entscheidung 2019/16 trifft ausschließlich Festlegungen zu IT-Verfahren, „die dem Bund-Länder übergreifenden Datenaustausch oder dem Datenaustausch mit Bürgern und Wirtschaft dienen“. Es steht den Mitgliedern des IT-Planungsrats selbstverständlich frei, auch alle anderen IT-Verfahren konform zur DIN SPEC 91379 zu gestalten, oder dies nicht zu tun, wenn fachliche oder wirtschaftliche Gründe dagegen sprechen. Dafür bedarf es keiner Beschlussfassung des IT-Planungsrats.

E.2.5. Der Betrieb eines IT-Verfahrens (bzw. der Datenaustausch mit einem IT-Verfahren) beruht auf einer rechtlichen Regelung, in der unter anderem auch der zu verwendende Zeichensatz festgelegt ist. Leider

ist dieser auf der Basis bestehender Regelungen zu nutzende Zeichensatz „kleiner“ als der der DIN SPEC 91379. Das heißt, in dem vom IT-Planungsrat vorgegebenen Zeichensatz gibt es Zeichen, die für das IT-Verfahren nicht zulässig sind. Was gilt denn jetzt?

Die Entscheidung 2019/16 des IT-Planungsrats kann entgegenstehende rechtliche Regelungen nicht aushebeln. Es ist deshalb erforderlich, solche Regelungen zu ändern. Es ist die Aufgabe der Mitglieder des IT-Planungsrats, also des Bundes und der Länder, im Wege der Umsetzung der Entscheidung 2019/16 im eigenen Verwaltungsraum solche Regelungen zu identifizieren und deren Änderung zu initiieren.

- E.2.6.** Sind auch solche IT-Verfahren betroffen, bei denen der Datenaustausch nur auf dem Papierweg erfolgt, zum Beispiel in Form von regelmäßigen Mitteilungen an Bürgerinnen und Bürger auf dem Postweg?

Die verbindliche Vorgabe gemäß Ziffer zwei der Entscheidung 2014/04 basiert auf der Kompetenz des IT-Planungsrats, fachunabhängige und fachübergreifende IT-Interoperabilitäts- und IT-Sicherheitsstandards zu beschließen. Die Vorgabe ist daher nur für solche IT-Verfahren verbindlich, bei denen der Datenaustausch in elektronischer Form erfolgt.

Beachten Sie aber, dass es einen Rechtsanspruch auf die korrekte Schreibweise des Namens in den Schreiben der öffentlichen Verwaltung an Bürgerinnen und Bürger gibt. Insofern sind Konstellationen denkbar, bei denen ein solcher Rechtsanspruch bisher mit Hinweis auf die nicht ausreichende technische Leistungsfähigkeit des jeweiligen IT-Verfahrens abgewehrt worden ist. Zukünftig wird es so sein, dass die technische Leistungsfähigkeit gegeben sein muss, sofern für das betreffende IT-Verfahren die Entscheidung 2014/04 des IT-Planungsrats einschlägig ist. Insoweit sind mittelbare Auswirkungen auch auf den nicht-elektronischen Datenaustausch zu erwarten.

- E.2.7.** Welche Auswirkungen haben die Entscheidungen 2014/04 und 2019/16 auf die Daten, die derzeit bereits in elektronisch geführten Registern gespeichert sind? Sie entsprechen dem derzeitigen Stand der Technik der IT-Verfahren. Müssen diese Registereinträge nachträglich angepasst werden, wenn die IT-Verfahren konform zur DIN SPEC 91379 sein werden?

Die Entscheidungen 2014/04 und 2019/16 des IT-Planungsrats erfordern dies nicht notwendigerweise. Die Konformität eines IT-Verfahrens ist hergestellt, wenn die Möglichkeit der vollumfänglichen Verarbeitung und Speicherung aller Zeichen der Spezifikation geschaffen worden ist. Inwiefern von dieser Möglichkeit Gebrauch gemacht wird, insbesondere ob sich die neuen Möglichkeiten auch auf Bestandsdaten und / oder archivierte Daten beziehen, ist für die Feststellung der Konformität irrelevant.

Die Frage muss daher unter fachlichen und wirtschaftlichen Aspekten jeweils im Einzelfall bewertet und entschieden werden. Die Herstellung der Konformität eines IT-Verfahrens zur Spezifikation kann dazu führen, dass auch Bestandsdaten bzw. archivierte Daten betroffen sind.

Zur Illustration des Sachverhalts und der möglichen Auswirkungen auf bereits vorhandene Registereinträge kann auf das Beispiel des Meldewesens verwiesen werden. Die dort genutzten IT-Verfahren verfügten vor der verbindlichen Vorgabe des Standards „Lateinische Zeichen in Unicode“ in der Regel nur über einen deutlich reduzierten Zeichensatz, so dass Namen von Personen in manchen Fällen in einer Schreibweise geführt wurden, die vom Eintrag in den Personenstandsregistern abwich. Die verbindliche Vorgabe des Standards „Lateinische Zeichen in Unicode“ führte nicht nur dazu, dass die Funktionalität der IT-Verfahren für zukünftige Einträge verbessert wurde. Aus fachlichen Gründen führte sie auch dazu, dass die bisher aus technischen Gründen in abweichender Schreibweise geführten Registereinträge modifiziert werden, so dass der in den Melderegistern gespeicherte Name identisch zum Eintrag im Personenstandsregister und zur Schreibweise in den entsprechenden hoheitlichen Dokumenten (Personalausweis, Pass) sein wird. Diese Angleichung der Schreibweise in Melderegistern kann aber nicht automatisiert erfolgen, sondern sie erfolgt anlassbezogen, z. B. bei der Beantragung eines neuen Personalausweises. Die Umstellung der Bestandsdaten wird daher bis zu zehn Jahre in Anspruch nehmen.

- E.2.8.** Möglicherweise sind Investitionen erforderlich, um die Konformität zur Spezifikation für IT-Verfahren herzustellen, für die die Entscheidung 2019/16 des IT-Planungsrats einschlägig ist. Wer trägt die Kosten?

Im Beschluss der Spezifikation sind keine speziellen Regelungen für Übernahme von Kosten enthalten. Daher muss mit den etwaigen Umsetzungskosten so umgegangen werden, wie das bei anderen Beschlüssen der IT-Steuerung innerhalb der jeweiligen Gebietskörperschaften üblich ist. Die Umsetzungsfristen (siehe Entscheidung 2019/53) sind so ausgestaltet, dass ausreichend Zeit für die Anmeldung der benötigten Haushaltsmittel in den Aufstellungsverfahren bleiben sollte.

E.3. Formulierungshilfen für rechtliche Vorgaben sowie Beschaffung von IT-Verfahren

- E.3.1.** Mit welcher Formulierung wird sichergestellt, dass ein IT-Verfahren alle normativen Zeichen der DIN SPEC vollständig verarbeiten kann?

Für diesen Zweck wird die Formulierung *"Das IT-Verfahren muss zur DIN SPEC 91379 konform sein"* vorgeschlagen.

Die Konformitätsbestimmung im Abschnitt 4.1 legt fest, dass ein IT-Verfahren dann konform zur DIN SPEC ist, wenn es alle normativen Zeichen bei allen für Namen (im weiteren Sinne) vorgesehenen Datenfeldern vollständig verarbeiten kann.

- E.3.2.** Mit welcher Formulierung wird sichergestellt, dass ein IT-Verfahren alle in der DIN SPEC aufgeführten Zeichen vollständig verarbeiten kann?

Für diesen Zweck wird die Formulierung *"Das IT-Verfahren muss zur DIN SPEC 91379 konform sein und darüber hinaus auch die erweiterten Schriftzeichen des Anhangs A der DIN SPEC 91379 vollständig unterstützen"* vorgeschlagen.

Durch die ausdrückliche Benennung auch der in Abschnitt A genannten Schriftzeichen wird klargestellt, dass die Konformität des IT-Verfahrens nicht ausreichend ist, sondern dass das IT-Verfahren auch kyrillische und griechische Zeichen vollständig unterstützen muss.

E.4. Formulierungshilfen für Schnittstellenbeschreibungen

- E.4.1.** Werden weiterhin Schnittstellenvereinbarungen benötigt oder reicht es aus, die DIN SPEC 91379 zu vereinbaren, um eine Schnittstelle zum Datenaustausch vollständig zu beschreiben?

Die DIN SPEC 91379 legt lediglich einen Zeichenvorrat fest. Dies reicht nicht aus, um die Syntax und Semantik der auszutauschenden Daten hinreichend festzulegen. Konkrete Schnittstellenvereinbarungen bleiben erforderlich.

Wenn es fachlich notwendig sein sollte, mehr Zeichen zu unterstützen, als in der Spezifikation aufgeführt werden, muss dies ebenfalls durch Schnittstellenvereinbarungen geregelt werden.

- E.4.2.** Mit welcher Formulierung wird festgelegt, dass eine Schnittstelle ausschließlich die von der DIN SPEC 91379 normativ vorgegebenen Zeichen unterstützt?

Die gesuchte Formulierung muss klar zum Ausdruck bringen, dass ausschließlich die in Abschnitt 4 aufgeführten Zeichen unterstützt werden, während die in Abschnitt A erweiterten Zeichen ebenso wenig unterstützt werden, wie sonstige Zeichen außerhalb der DIN SPEC.

Darüber hinaus wird empfohlen, die Unicode Normalform C (NFC) zu fordern. Dies gilt insbesondere dann, wenn die informell bereitgestellten Umsetzungen der in Anhang 9 genannten Datentypen zum Einsatz kommen. Zur Begründung wird auf Abschnitt B.7 dieses Dokuments verwiesen.

Für diesen Zweck wird folgende Formulierung vorgeschlagen: *"Für den Datenaustausch wird der durch die DIN SPEC 91379 normativ bestimmte Zeichensatz festgelegt. Die Daten müssen gemäß der Unicode Normalform C (NFC) normalisiert sein."*

Die Formulierung *"Die Schnittstelle ist konform zur DIN SPEC 91379"* wird *nicht empfohlen*. Zwar wird damit - in analoger Anwendung des Konformitätsbegriffs aus Abschnitt 4.1 der DIN SPEC - sichergestellt, dass jedes durch die DIN SPEC 91379 normativ vorgegebene Zeichen unterstützt wird. Die Formulierung lässt aber die Frage offen, ob darüber hinaus gehende Zeichen, zum Beispiel solche aus dem erweiterten Bereichen in Anhang A der DIN SPEC, ebenfalls unterstützt werden, oder ob sie als fehlerhaft zurückgewiesen werden dürfen.

- E.4.3.** Mit welcher Formulierung wird festgelegt, dass eine Schnittstelle sowohl die normativen, als auch die empfohlenen Zeichen der DIN SPEC 91379 unterstützt?

Für diesen Zweck wird folgende Formulierung vorgeschlagen: *"Für den Datenaustausch wird der durch die DIN SPEC 91379 bestimmte Zeichensatz inklusive der erweiterten Zeichen des Anhangs A festgelegt. Die Daten müssen gemäß der Unicode Normalform C (NFC) normalisiert sein."*

Die ausdrückliche Nennung des erweiterten Zeichensatzes in Anhang A der DIN SPEC hilft, Missverständnisse zu vermeiden.

E.5. Umfang der DIN SPEC 91379

- E.5.1.** Warum wurden Zeichen aus weit verbreiteten Zeichensätzen nicht in die Spezifikation übernommen? Die DIN SPEC 91379 ist deshalb z. B. nicht kompatibel zu ISO 8859-1, ISO 8859-15 und CP1252.

Betroffen sind ca. 16 Zeichen (ohne Steuerzeichen) aus der ISO-8859- und der CP1252-Familie, also aus allen ISO 8859-* Normen und allen Windows-Standards und nicht nur ISO 8859-1, ISO 8859-15 und CP1252.

Im Abschnitt F wird erläutert, warum Zeichen aus den drei genannten Standards nicht in die Spezifikation übernommen wurden. Allgemein lässt sich sagen, dass die fehlenden Zeichen keine Relevanz für die Darstellung von Namen haben. Eine Kompatibilität zu den genannten Standards wird bereits allein aufgrund der Entscheidung, auf Steuerzeichen aus der Zeit von Lochstreifen und Fernschreiber zu verzichten, von vornherein ausgeschlossen.

- E.5.2.** Wie wurde die Menge der lateinischen Zeichen bestimmt?

Die Festlegung des Zeichensatzes erfolgte iterativ. Der Ausgangspunkt war eine einschlägige Dissertation, in der ein „paneuropäischer Zeichensatz“ auf Basis der Amtssprachen der europäischen Mitgliedsstaaten vorgeschlagen worden ist.

Dieser erste Vorschlag wurde zunächst im Innenressort, anschließend mit weiteren Ressorts abgestimmt. Die resultierende Version 1.1.1 wurde zunächst in Teilen der Innenverwaltung (Ausländer-, Melde- und Personenstandswesen), anschließend durch den IT-Planungsrat beschlossen.

Die Version 1.1 von String.Latin enthält

- Alle lateinischen Schriftzeichen der EU-Amtssprachen sowie der Amtssprachen Islands, Liechtensteins, Norwegens und der Schweiz,
- Alle Schriftzeichen der deutschen Minderheitensprachen,
- Schriftzeichen, die von Personenstandsregistern in Deutschland für Namen natürlicher Personen benötigt werden.

Im Rahmen der XÖV-Konferenz 2016 wurde eine Fachgruppe String.Latin ins Leben gerufen. Die Fachgruppe setzt sich zusammen aus Behörden und Dienstleistern, die mit der Einführung und Anwendung des Zeichensatzes in der Praxis befasst sind. Sie berät die KoSIT und damit den IT-Planungsrat hinsichtlich der Umsetzung und der Weiterentwicklung der Spezifikation. Die Ausarbeitung einer neuen Version von String.Latin erfolgte ab 2018 in einem DIN-Workshop zur Erstellung der DIN SPEC 91379.

Alle Schriftzeichen aus String.Latin 1.1 mussten auch in der DIN SPEC 91379 normativ enthalten sein, um die Abwärtskompatibilität sicherzustellen. Darüber hinaus gab es Bedarf, zusätzliche Schriftzeichen aufzunehmen. Schriftzeichen, die Standesbeamte in String.Latin 1.1 für das Führen von Personenstandsregistern vermissten, wurden im normativen Teil ergänzt. Außerdem kamen neu hinzu:

- Normativ alle Schriftzeichen, die das Ergebnis einer ISO-Transliteration sind, die also beim Übertragen von Namen anderer Schriftsysteme in die lateinische Schrift entstehen können,
- Nicht normativ Schriftzeichen, die für Namen juristischer Personen in Deutschland benötigt werden, die sich also bereits im Handelsregister befinden – dazu gehören neben lateinischen auch griechische Schriftzeichen,
- Nicht normativ griechische und kyrillische Schriftzeichen, um für den grenzüberschreitenden Datenaustausch auch die Amtssprachen von Griechenland, Zypern und Bulgarien abzudecken. Dazu wurden die griechischen und kyrillischen Zeichen aus dem SEPA-Zeichensatz übernommen, mit dem Banken ihren internationalen Datenaustausch regeln.

Anders ausgedrückt fehlen in der DIN SPEC 91379 lateinische Schriftzeichen aus Minderheitensprachen anderer EU-Staaten sowie aus Amtssprachen und Minderheitensprachen von Staaten außerhalb der EU. Vereinzelt sind jedoch auch solche Zeichen bereits enthalten, wenn sie in deutschen Personenstandsregistern vorkommen.

E.5.3. Auf welcher Basis wurde der Zeichenumfang für die Datentypen festgelegt für

- *Namen natürlicher Personen,*
- *Anschriften,*
- *Namen juristischer Personen und*
- *Produktnamen?*

Grundsätzlich sei zur Erläuterung der fünf Datentypen A bis E auf die Ausführungen im Abschnitt B verwiesen. Die Zuordnung der Schriftzeichen in die jeweiligen Datentypen erfolgte im Wesentlichen auf der Basis der Anforderungen der datenverarbeitenden Stellen. Als die wichtigsten sind zu nennen:

- *Das Personenstandswesen für Namen natürlicher Personen,*
- *Das Meldewesen für deutsche Anschriften,*
- *Das Ausländerwesen für ausländische Anschriften,*
- *Registergerichte bzw. Handelsregister für Namen juristischer Personen,*
- *Das Deutsche Patent- und Markenamt (DPMA) für Produktnamen,*
- *Das Statistische Bundesamt im Auftrag des Auswärtigen Amtes für Namen von Staaten.*

E.5.4. Welche Schriftzeichen sind in der Spezifikation enthalten?

Die DIN SPEC 91379 listet alle normativen Schriftzeichen in Tabelle 3 bis Tabelle 8 auf. In Tabelle A.1 bis Tabelle A.3 werden die nicht-normativen Schriftzeichen aufgeführt.

In den Tabellen sind insgesamt 924 Zeichen enthalten, die meisten davon sind Buchstaben. Neben den lateinischen Grundbuchstaben A...Z und Eszett (ß) aus dem deutschen Alphabet gibt es weitere Grundbuchstaben, wie z. B. die Ligaturen Æ, Œ und IJ, die Zeichen Thorn (Þ), Kra (κ), Eng (Ŋ), Schwa (Θ) und Ezh (ʒ) sowie das türkische I ohne Punkt (ı). Die meisten dieser Grundbuchstaben können jeweils als Groß- und Kleinbuchstabe auftreten und mit diakritischen Zeichen kombiniert sein. Hinzu kommen:

- Griechische und kyrillische Buchstaben
- Die Ziffern 0...9 und weitere Zahlzeichen, wie z. B. $\frac{1}{2}$, tief- und hochgestellte Ziffern;
- Hochgestellte Buchstaben, um englische Ordinalzahlen darstellen zu können, z. B. 1st;
- Interpunktionszeichen, wie z. B. öffnende und schließende Klammern, Frage- und Ausrufezeichen und so weiter;
- Das Leerzeichen und diverse sonstige Nicht-Buchstaben.

Zum Vergleich: viele bisher genutzte Zeichensätze (Legacy Codepages) wie z. B. ISO 8859-1 und ISO 8859-15 verfügen aus technischen Gründen nur über maximal 256 Zeichen. In dem Unicode-Standard Version 12 [Unicode 12] sind etwa 138.000 Zeichen enthalten. Hinzu kommen die diversen flexiblen Kombinationsmöglichkeiten mit diakritischen Zeichen. Die Spezifikation enthält also deutlich mehr Zeichen, als von Legacy Codepages unterstützt werden, aber auch deutlich weniger Zeichen als in Unicode enthalten sind.

E.5.5. Bleiben einmal aufgenommene Zeichen in zukünftigen Fassungen der Spezifikation erhalten?

Die Regeln zur Pflege der Spezifikation müssen noch festgelegt werden. Da während der Weiterentwicklung der Publikationstyp gewechselt wird (KoSIT-Publikation, DIN SPEC, DIN-Norm, CEN-Norm), können noch keine verbindlichen Aussagen getroffen werden. Es ist aber davon auszugehen, dass ein Zeichen, welches einmal in den Standard „Lateinische Zeichen in Unicode“ (String.Latin 1.1) oder in die DIN SPEC 91379 (String.Latin+ 1.2) aufgenommen worden ist, in zukünftigen Fassungen der Spezifikation erhalten bleiben muss. Denn die Spezifikation ist eine verbindliche Vorgabe nicht nur für die Datenübermittlung, sondern auch für die Registerführung. Das Entfernen einzelner Zeichen aus der Spezifikation im Rahmen der Fortschreibung würde dazu führen, dass Registereinträge, die derzeit nur in der Spezifikation aufgeführte normative Schriftzeichen enthalten, zukünftig Probleme im Datenaustausch mit anderen Systemen bekommen könnten, wenn einzelne Zeichen ihre Verbindlichkeit verlieren. Dies gilt es zu vermeiden.

Auch für den von einem Industriekonsortium geführten Standard Unicode gilt die Policy, dass ein einmal aufgenommenes Zeichen in allen zukünftigen Fassungen erhalten bleiben muss. Es wird angestrebt, die gleiche Regel auch für die Weiterentwicklung der DIN SPEC einzuhalten.

String.Latin+ 1.2 ist jedenfalls abwärtskompatibel zu String.Latin 1.1, d.h. dass alle Zeichen, die in „Lateinische Zeichen in Unicode“ enthalten waren, sich im normativen Teil der DIN SPEC 91379 befinden.

E.5.6. Durch die Spezifikation wird eine bestimmte Menge von Zeichen abschließend aufgezählt. Was passiert, wenn ich aus fachlichen Gründen Zeichen in einem IT-Verfahren benötige, die nicht in der Spezifikation genannt sind?

Die Konformität eines IT-Verfahrens ist in dem Moment gegeben, in dem alle in der Spezifikation genannten Zeichen voll unterstützt werden. Die Unterstützung darüber hinausgehender Zeichen, die in der Spezifikation nicht genannt werden, steht der Konformität nicht entgegen und ist insoweit völlig unproblematisch. Dies ist der Grund, warum der Standard „Lateinische Zeichen in Unicode“ (String.Latin 1.1) in der Entscheidung 2014/04 als „Mindeststandard“ bezeichnet wird. Auch die DIN SPEC 91379 (String.Latin+ 1.2) legt die Schriftzeichen im normativen Teil als „Mindestmenge“ fest (siehe Kapitel 1). Weiter heißt es „Zur DIN SPEC konforme IT-Verfahren dürfen darüber hinausgehende Schriftzeichen unterstützen.“

Allerdings können Sie sich nicht darauf verlassen, dass andere IT-Verfahren mit diesen, über die Spezifikation hinausgehenden Zeichen, ebenfalls umgehen können. Insofern sind ggf. zusätzliche Schnittstellenvereinbarungen erforderlich, um zu klären, ob die zusätzlichen Zeichen auch problemlos an andere IT-Verfahren übermittelt und dort verarbeitet werden können.

Wenn Sie den Eindruck haben, dass das bzw. die von Ihnen benötigte Zeichen in die Spezifikation aufgenommen werden sollten, damit sie zukünftig von allen wichtigen IT-Verfahren der öffentlichen Verwaltung unterstützt werden, dann richten Sie bitte einen Änderungsantrag an die KoSIT.

E.5.7. Warum gibt es im Abschnitt D Lücken, also Zeichen der DIN SPEC 91379, die nicht auf Legacy-Zeichensätze abgebildet werden?

Im Generellen wurde beim Mapping versucht, die Semantik und nicht das Erscheinungsbild eines Zeichens beizubehalten. Deshalb kann es passieren, dass kein geeignetes Mapping gefunden wurde.

„Lücken“ finden sich in der aktuellen Fassung des Mappings bei fünf Zeichen, welche thematisch in zwei Gruppen aufgeteilt werden müssen:

- Die Zeichen U+00A6 BROKEN BAR (|), U+00A8 DIAERESIS (¨), U+00B4 ACUTE ACCENT (´) und U+00B8 CEDILLA (¸) könnten zwar syntaktisch ähnlich repräsentiert werden (z.B. durch | ´ ¸), würden hierdurch aber ihre gesamte Semantik verlieren und somit den Sinn der Zeichenkette verändern oder gar verfälschen.
- Das Zeichen U+00A4 CURRENCY SIGN (¤) auf ein (z.B. lokales) Währungszeichen zu mappen, kann vor allem beim Datenaustausch von Rechnungen und Geldbeträgen sehr gefährlich sein. Sollte es bei Datenaustauschpartner A noch in Währung X vorhanden gewesen sein und durch unaufmerksames Mapping bei Datenaustauschpartner B in Währung Y ankommen, lässt sich der wirtschaftliche Verlust leicht errahnen.

E.6. Umgang mit unterschiedlichen Zeichensätzen in IT-Verfahren

E.6.1. Kann man sich darauf verlassen, dass alle IT-Verfahren, die konform zur DIN SPEC 91379 sind, exakt den gleichen Zeichensatz unterstützen?

Nein. Jedes zur Spezifikation konforme IT-Verfahren muss mindestens die in der Spezifikation genannten Zeichen unterstützen. Aber es kann darüber hinaus andere Zeichen unterstützen, zum Beispiel aufgrund entsprechender fachlicher Anforderungen.

Insofern kann man sich bei Datenübermittlungen nicht darauf verlassen, dass von IT-Verfahren, die konform zur Spezifikation sind, ausschließlich Zeichen der Spezifikation übermittelt werden. Je zwei zur DIN SPEC konforme IT-Verfahren können sehr unterschiedliche Zeichensätze unterstützen, aber sie verfügen zumindest über eine gemeinsame, für „normale Zwecke“ ausreichende Schnittmenge. Präzise Festlegungen zu Schnittstellen, inklusive des jeweils zu verwendenden Zeichensatzes, bleiben auch weiterhin erforderlich.

E.6.2. Mein eigenes IT-Verfahren ist konform zur DIN SPEC 91379. Ich bin zum Datenaustausch mit einem anderen IT-Verfahren verpflichtet, welches lediglich einen 8-Bit-Zeichensatz (beispielsweise ISO 8859-15) oder String.Latin 1.1 unterstützt. Bei einer Konvertierung meiner Daten in den weniger leistungsfähigen Zeichensatz entstehen Informationsverluste. Zudem entstehen in meiner eigenen Organisationseinheit Kosten, deren Verursacher der Betreiber mit dem weniger leistungsfähigen Zeichensatz ist. Was ist zu tun?

Falls das andere IT-Verfahren von einer deutschen Behörde betrieben wird und lediglich einen 8-Bit-Zeichensatz unterstützt, wenden Sie sich mit ihrem Anliegen an die Geschäftsstelle des IT-Planungsrats. Gemäß Entscheidung 2014/04 müssen IT-Verfahren der öffentlichen Verwaltung, die dem bund-länder-übergreifenden Datenaustausch oder dem Datenaustausch mit Bürgern und Wirtschaft dienen, seit März 2017 den Standard „Lateinische Zeichen in Unicode“ (String.Latin 1.1) unterstützen.

Insofern muss von Seiten der zuständigen Gremien des IT-Planungsrats dem Verdacht nachgegangen werden, dass ein Verstoß gegen die Beschlusslage des IT-Planungsrats vorliegt. Das weitere Vorgehen hängt vom Ergebnis dieser Prüfung ab und muss im Einzelfall festgelegt werden.

Falls das andere IT-Verfahren von einer deutschen Behörde betrieben wird und String.Latin 1.1 unterstützt, besteht gemäß der Entscheidung 2019/53 des IT-Planungsrats eine Übergangsfrist bis zum 1. November 2024, um die DIN SPEC 91379 zu unterstützen. Um die Informationsverluste zu minimieren, sollten nur die Zeichen, welche ausschließlich in der DIN SPEC 91379 enthalten sind, jedoch nicht in String.Latin 1.1, einem Mapping unterworfen werden. Dabei kann sich an der Abbildung auf Legacy Codepages (z. B. auf ISO 8859-15) orientiert werden, siehe D .

- E.6.3.** Ist die Suchform (Abschnitt 5 „Normative Abbildung lateinischer Buchstaben auf Grundbuchstaben (Suchform)“) eine Alternative zu den Legacy-Mappings (Abschnitt 9 „Abbildung auf Legacy Codepages“)?

Nein. Die Suchform kann man als Mapping in den ICAO-Zeichensatz verstehen, den sehr eingeschränkten Zeichensatz, der für den maschinenlesbaren Teil von Ausweisen und Pässen zum Einsatz kommt. Die Legacy-Mappings schlagen jeweils eine konkrete Abbildung in die Zeichensätze ISO 8859-1, ISO 8859-15, Windows CP1252 sowie EBCDIC CP1141 vor.

- E.6.4.** Aufgrund internationaler Vorgaben wird mir vorgeschrieben, wie mein IT-Verfahren lateinische Buchstaben in einer phonetischen Suche zu behandeln hat. In dem internationalen Standard (z. B. NYSIS-Phonschlüssel) sind aber nicht alle Zeichen aus String.Latin+ enthalten. Was soll ich tun?

String.Latin+ setzt die Rechtslage in Deutschland und aller weiteren Staaten um, die sich dem Übereinkommen über die Angabe von Familiennamen und Vornamen in den Personenstandsbüchern vom 13. September 1973 angeschlossen haben (siehe Abschnitt *Einleitung*). Wenn die internationalen Vorgaben auch auf Namen angewendet werden sollen, dann ist es empfehlenswert, auf eine Erweiterung dieser Vorgaben hinzuwirken. Wenn Zeichen aus String.Latin+ nicht unterstützt werden, sind die Vorgaben lückenhaft.

Da Änderungen internationaler Vorgaben i.d.R. sehr langwierig sind, könnte es zur Überbrückung helfen, eine der Abbildungen auf Legacy-Zeichensätze (siehe Abschnitt D) zu verwenden und die internationalen Vorgaben auf das Ergebnis der Abbildung anzuwenden. Es ist zu erwarten, dass internationale Vorgaben Zeichensätze wie ISO 8859-1 oder ISO 8859-15 vollständig unterstützen.

Falls die internationalen Vorgaben gar nicht geändert werden können, könnte versucht werden, die Spezifikation String.Latin+ um vollständige Vorgaben zu erweitern. Das Ziel der Standardisierung ist eine CEN-Norm. Eine vollständige Regelung in einer Norm könnte die internationalen Vorgaben in der Zukunft ersetzen. Dafür wird allerdings die Beteiligung entsprechender Fachexpert/innen (im Fall von NYSIS-Phonschlüssel Expert/innen für phonetische Suchen) im Standardisierungsprozess notwendig sein.

E.7. Implementation der DIN SPEC 91379

- E.7.1.** Ist Unicode-Fähigkeit von Schnittstellen ausreichend?

Nein, das reicht nicht. Solche Vorgaben regeln üblicherweise den Datenaustausch, nicht aber die Leistungsfähigkeit der im Verbund beteiligten IT-Verfahren. Natürlich kann man aus Vorgaben zum Datenaustausch zumindest indirekt Rückschlüsse über die als Sender bzw. Empfänger beteiligten Stellen ziehen. Man darf erwarten, dass IT-Verfahren die Nachrichten, die gemäß dieser Vorgaben übermittelt werden, auch angemessen verarbeiten. Ohne weitere Vorgaben ist aber nicht sichergestellt, dass diese Verarbeitung verlustfrei erfolgen muss. Zeichen, die nicht direkt verarbeitet werden können, werden ggf. in eine „möglichst ähnliche“ Darstellung überführt oder fehlerhaft dargestellt.

Im Unterschied dazu trifft die DIN SPEC 91379 im 4.1 Aussagen zu geforderten Eigenschaften von IT-Verfahren. Jedes Zeichen der Spezifikation muss vollumfänglich verarbeitet werden können. „Dies umfasst insbesondere die Erfassung, Speicherung, Übermittlung, Anzeige und den Ausdruck aller Schriftzeichen.“

Außerdem wird durch die Formulierung „Teilmenge der in Unicode enthaltenen Zeichen und Zeichensequenzen“ festgelegt, dass Unicode nicht in Gänze unterstützt werden muss (was nach heutigem Stand der Technik auf der Anwendungsebene unrealistisch ist).

E.7.2. Ist die DIN SPEC 91379 korrekt implementiert, wenn ein IT-Verfahren Unicode vollständig unterstützt?

Grundsätzlich lautet die Antwort: ja. Wenn alle Unicode-Zeichen unterstützt werden, werden auch alle Zeichen der DIN SPEC 91379 unterstützt. Nachzuweisen, dass die Erfassung, Speicherung, Übermittlung, Anzeige und der Ausdruck aller Unicode-Zeichen korrekt sind (siehe Abschnitt 4.1), gestaltet sich angesichts der Vielzahl an Zeichen und Kombinationsmöglichkeiten jedoch schwierig. Seit Unicode 10 enthält der Standard allein 130.000 Zeichen. Hinzu kommen die unzähligen Kombinationsmöglichkeiten, mit denen Zeichensequenzen gebildet werden können. Zur korrekten Implementation der DIN SPEC gehört die Unterstützung von 606 normativen Zeichen und 147 normativen Zeichensequenzen (davon 11 mit drei Codepoints). Auch bei einer theoretischen Unterstützung aller Unicode-Zeichen sollte die korrekte Erfassung, Speicherung, Übermittlung, Anzeige und Ausdruck der 753 normativen Zeichen und Zeichensequenzen überprüft werden.

Aus fachlichen Gründen kann es dringend geboten sein, die Menge aller Unicode-Zeichen einzuschränken, siehe dazu den Abschnitt „Abwägung von Alternativen“. Als Beispiel sei genannt, dass die lateinische Zeichenkette „Paul“ anstelle des lateinischen P auch mit einem griechischen RHO („Paul“) oder einem kyrillischen ER („Paul“) geschrieben werden kann. Die Namen sind informationstechnisch unterschiedlich – für das menschliche Auge sind die drei Glyphen identisch. Eine korrekte Identifikation von Personen wird dadurch erschwert.

Unicode bietet vor allem durch die Kombinationsmöglichkeiten von Zeichen noch viel mehr Potenzial, unterschiedliche Schriftzeichen zu erzeugen, die mit einer identischen oder sehr ähnlichen Glyphen repräsentiert werden.

Die Einschränkung der zulässigen Zeichen auf die Datentypen A bis D aus dem Abschnitt B zusammen mit der Verwendung der Unicode Normalform C (NFC) beugt den meisten missbräuchlichen Einsatzmöglichkeiten von Unicode vor.

E.7.3. Enthält der Datentyp C wirklich alle normativen Zeichen? Laut „Tabelle B.1: Übersicht über die empfohlenen Datentypen und zugehörigen Schriftzeichengruppen“ und Abschnitt 7.4 „Datentyp C“ scheinen die normativen Zeichen der Tabelle 4 zu fehlen.

Ja, im Datentyp C sind alle normativen Zeichen der DIN SPEC 91379 enthalten, auch die kombinierenden diakritischen Zeichen der Tabelle 4. Diese sind jedoch immer ein Bestandteil abschließend festgelegter Kombinationen in Zeichensequenzen. Sie dürfen nicht alleinstehend im Datentyp C vorkommen, weil dann beliebige Kombinationen mit diesen Zeichen gebildet werden könnten.

Zur näheren Erklärung: Die Tabelle 3 enthält alle lateinischen Buchstaben. Einige der lateinischen Buchstaben von Tabelle 3 sind kodierte Zeichensequenzen (siehe Begriffsdefinition 3.4 „kodierte Zeichensequenz“). Sie bestehen aus einem Grundbuchstaben sowie einem oder mehreren der in Tabelle 4 genannten „kombinierenden diakritische Zeichen“. Diese bedürfen einer besonderen Betrachtung, denn sie verändern die Bedeutung des Grundbuchstabens, mit dem sie kombiniert werden.

- Einerseits muss jedes IT-Verfahren in der Lage sein, die Zeichen in Tabelle 4 zu verarbeiten. Wäre dies nicht der Fall, dann könnte das IT-Verfahren keine der in Tabelle 3 enthaltenen Zeichensequenzen darstellen. Das ist der Grund, warum Tabelle 4 zu den normativen Zeichen gehört.
- Andererseits soll mit den Datentypen verhindert werden, dass andere als die in Tabelle 3 enthaltenen Zeichensequenzen übermittelt werden können. Das wäre aber der Fall, wenn die Zeichen der Tabelle wie alle anderen Zeichen beliebig übermittelt werden können. Man könnte dann beispielsweise ein großes ß mit kombinierendem CIRCUMFLEX und OGONEK übermitteln – das ist fachlicher Unfug und soll verhindert werden.
- Deswegen heißt es in Abschnitt B.1, fünfter Absatz, vierter Satz: „Die regulären Ausdrücke setzen zudem die Anforderung um, dass kombinierende diakritische Zeichen ausschließlich in den Zeichensequenzen genutzt werden, die in Tabelle 3 abschließend genannt sind.“ Die Zeichen der Tabelle sind also in dem Datentyp C bereits enthalten, allerdings immer nur als Bestandteil einer Zeichensequenz aus Tabelle 3.

E.7.4. Wie werden die Datentypen mithilfe der mitgelieferten XML-Anhänge implementiert?

Jede moderne Programmiersprache und jedes Rahmenwerk für Datenübermittlungen bietet grundlegende Datentypen für Zeichenketten, die auf Unicode basieren. In XML ist dies der Datentyp *xs:string*. Mit diesem Datentypen lassen sich nahezu alle Zeichen des Unicode-Standards verarbeiten bzw. übermitteln.

Für die technischen Datentypen des Abschnitts 7 werden Implementationen in XML als digitale Anhänge zur DIN SPEC mit ausgeliefert, die den Datentyp *xs:string* auf die in den jeweiligen Datentypen enthaltenen Zeichen einschränken. Die Nutzung eines Zeichens außerhalb der Spezifikation führt bei der Nutzung dieser XML-Datentypen zu einem technischen Fehler.

Die Datentypen können insoweit die Realisierung von XML-basierten Schnittstellen erleichtern. Sie entlasten die am Datenaustausch beteiligten IT-Verfahren von der Aufgabe der Prüfung, ob Zeichen genutzt werden, die in der Spezifikation nicht aufgeführt sind.

Ursprünglich wurde von der KoSIT der XML-Datentyp „String.Latin“ entwickelt. Er enthält alle Zeichen des Standards „Lateinische Zeichen in Unicode Version 1.1.1“ (String.Latin 1.1) und erleichterte die Verwendung des Zeichenvorrats in XÖV-Standards.

E.7.5. Welche Schriftarten stellen alle Unicode-Zeichen der DIN SPEC 91379 korrekt dar?

Verwenden Sie eine lizenzgebundene Schriftart für Ihre Anwendung, klären Sie bitte mit Ihrem Fontanbieter die Verfügbarkeit aller Zeichen oder gegebenenfalls eine Erweiterung des Fonts im Rahmen Ihrer Lizenz (unter Windows ist z.B. die Schriftart „Calibri Light“ verfügbar, mit der gute Erfahrungen gemacht und noch keine falschen Darstellungen entdeckt wurden).

Es sind kostenfreie Schriftarten verfügbar, die den Anspruch verfolgen, alle Unicode-Zeichen korrekt darzustellen. Bei Wikipedia ist eine Übersicht erhältlich, siehe [Open Source Fonts]. Aus einer Analyse im März 2019 der gängigen Open Source Fonts [Liberation Fonts], [Dejavu Fonts] und [Noto Fonts], welche nicht nur frei verfügbar, sondern auch eigenständig erweitert oder geändert werden können, ging die Kompatibilitätsmatrix in der Tabelle E.1 hervor. In Klammern wird die Anzahl Glyphen genannt, die mit der jeweiligen Schriftart nicht korrekt darstellbar sind.

Je nach Anwendungsfall und notwendigem Erscheinen des Fonts kann eine Anwendung auf verschiedene Optionen zurückgreifen und im Zweifelsfall mit überschaubarem Aufwand eigene Zeichen hinzufügen, oder Änderungsanträge bei den Entwicklern einbringen.

Für weitere Informationen zu DejaVu, Liberation und Google Noto, siehe die dazugehörigen Links.

E.7.6. Wird die Schriftart UnicodeDoc, die alle Zeichen des Standards „Lateinische Zeichen in Unicode“ (String.Latin 1.1) unterstützt, weiterentwickelt, um alle Zeichen der DIN SPEC 91379 (String.Latin+ 1.2) zu unterstützen?

Nein, die öffentliche Verwaltung beabsichtigt nicht, als Entwickler einer Schriftart für die Zeichen der DIN SPEC 91379 tätig zu werden, das heisst, dass UnicodeDoc voraussichtlich nicht weiterentwickelt wird. Aufgrund der stark gestiegenen Bedeutung von Unicode für weltweite IT-Verfahren sind inzwischen kostenfreie Schriftarten verfügbar, die den Anspruch verfolgen, alle Unicode-Zeichen korrekt darzustellen. In [Noto Fonts] ist eine Übersicht verfügbar.

Tabelle E.1. Kompatibilitätsmatrix zwischen Schriftarten und Schriftzeichengruppen

<i>String.Latin+</i>	<i>Liberation Sans</i>	<i>Liberation Serif</i>	<i>Liberation Mono</i>	<i>DejaVu Sans</i>	<i>DejaVu Serif</i>	<i>DejaVu Sans Mono</i>	<i>Noto Sans (Phase III)</i>	<i>Noto Serif</i>	<i>Noto Mono</i>	<i>Noto Sans Mono (Phase III)</i>



<i>Latein. Buchsta- ben (nor- mativ) (0041 - 1EF9) 499 Gly- phen</i>	✓	✓	✓	✓	✓	☹ (53)	✓	✓	☹ (147)	✓
<i>Nicht- Buchsta- ben N1 (norma- tiv) (0020 - 2021) 18 Glyphen</i>	✓	✓	✓	✓	✓	☹ (1)	✓	✓	☹ (8)	✓
<i>Nicht- Buchsta- ben N2 (norma- tiv) (0021 - 20AC) 60 Glyphen</i>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
<i>Nicht- Buchsta- ben N3 (norma- tiv) (00A4 - 00BE) 6 Glyphen</i>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
<i>Nicht- Buchsta- ben N4 (norma- tiv) (0009 - 00A0) 4 Glyphen</i>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
<i>Griech. Buchsta- ben (erwei- tert) (0386 - 03CE) 69 Glyphen</i>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
<i>Kyrill. Buchsta- ben (erwei- tert)</i>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

(040D - 045D) 62 Glyphen										
Nicht-Buchstaben E1 (erweitert) (0192 - 2265) 40 Glyphen	☹ (13)	☹ (13)	☹ (17)	✓	✓	✓	✓	☹ (3)	☹ (22)	✓

E.7.7. Wie kann die Eingabe der vielen Zeichen der DIN SPEC implementiert werden? Wird es eine zentrale Hilfestellung für die Implementation geben?

Eine kostenfreie Lösung, welche explizit in Hinsicht auf die Eingabe von String.Latin(+) Zeichen entwickelt wurde, gibt es nach aktuellem Stand leider (noch) nicht.

Wenn Sie mit Windows arbeiten, haben Sie verschiedene Möglichkeiten, Sonderzeichen zu erfassen:

- Sonderzeichen erfassen mit der Bildschirmtastatur
 1. Auf der Tastatur: Auf die Windows-Taste  drücken.
 2. Bildschirmtastatur erfassen.
 3. In der Trefferliste: Auf **Bildschirmtastatur** klicken.
 4. Das Fenster **Bildschirmtastatur** wird geöffnet.
 5. An die Stelle klicken, an der Sie das Zeichen erfassen wollen.
 6. Im Fenster **Bildschirmtastatur**: Auf das Zeichen klicken, das Sie erfassen wollen.
 7. Das Zeichen wird eingefügt an der Stelle, an der der Cursor steht.
- Sonderzeichen erfassen mit der Windows Zeichentabelle
 1. Auf der Tastatur: Auf die Windows-Taste  drücken.
 2. Zeichen erfassen.
 3. In der Trefferliste: Auf **Zeichentabelle** klicken.
 4. Das Fenster **Zeichentabelle** wird geöffnet.
 5. Im Fenster **Zeichentabelle**: Auf das gewünschte Zeichen doppelklicken.
 6. Gewählte Zeichen werden im Feld **Zeichenauswahl** angezeigt.
 7. Das Zeichen per Copy-Paste in das Zielfeld übertragen.
- Sonderzeichen erfassen mit ALT-Codes

Sie können in Windows Sonderzeichen erfassen mit Hilfe von Tastenkombinationen, die die ALT-Taste beinhalten. Diese Tastenkombinationen werden ALT-Codes genannt.

8. An die Stelle klicken, an der Sie das Zeichen erfassen wollen.

9. ALT-Taste drücken und gedrückt halten.

1 ALT-Code (Zahlenfolge) wählen auf dem **Nummernblock** der Tastatur.

0.

Die ALT-Codes aus der Codepage 850 entnehmen Sie bitte dem entsprechenden Eintrag bei Wikipedia [https://de.wikipedia.org/wiki/Codepage_850] oder – sofern Sie überdurchschnittlich gute Augen oder eine gute Lupe besitzen – der nachfolgenden Tabelle (numerischer Code in fetter, kursiver Schreibweise). Die Sonderzeichen, die zur Erfassung von Namen verwendet werden, sind grün hervorgehoben.

	_0	_1	_2	_3	_4	_5	_6	_7	_8	_9	_A	_B	_C	_D	_E	_F
0_	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1_	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
2_	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
3_	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4_	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5_	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7_	p	q	r	s	t	u	v	w	x	y	z	()	~	△	▽
8_	Ç	ü	é	à	ä	â	á	ç	ê	ë	è	í	ï	ì	Ä	Å
9_	É	æ	Æ	ô	ö	õ	ù	ý	Ö	Ü	ø	£	Ø	×	ƒ	
A_	á	í	ó	ú	ñ	Ñ	ª	º	¿	®	¬	½	¼	¡	«	»
B_	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
C_	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
D_	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
E_	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
F_	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255
	_0	_1	_2	_3	_4	_5	_6	_7	_8	_9	_A	_B	_C	_D	_E	_F

Bild 3: ALT-Codes zum Erfassen von Sonderzeichen

1 ALT-Taste loslassen. Das Zeichen wird eingefügt an der Stelle, an der der Cursor steht.

1.

Um die Bildschirmtastatur in einer anderen Sprache zu verwenden und dadurch ein anderes Tastaturlayout mit anderen Zeichen zu erhalten, ändern Sie die Eingabesprache in den Systemeinstellungen.

Wenn Sie mit macOS arbeiten, haben Sie verschiedene Möglichkeiten, Sonderzeichen zu erfassen:

- Sonderzeichen erfassen mit dem Fenster „Zeichen“

1. In der Menüleiste (standardmäßig oben rechts): Auf das Eingabemenü klicken.
2. Wenn das Eingabemenü nicht angezeigt wird oder keine Sprache angezeigt wird, muss zuerst eine weitere Sprache hinzugefügt werden.
3. Eintrag **Emoji & Symbole einblenden** wählen.
4. Das Fenster **Zeichen** wird geöffnet. In der linken Spalte sind Zeichen in Gruppen zusammengefasst. Die meisten in verwendbaren Zeichen sind in der Gruppe **Lateinisch** enthalten.
5. An die Stelle klicken, an der Sie das Zeichen erfassen wollen.
6. Im Fenster **Zeichen**: Auf das Zeichen doppelklicken, das Sie erfassen wollen.

- Sonderzeichen erfassen mit der Unicode Hex-Eingabe

1. Im Eingabemenü: Eintrag **Unicode Hex-Eingabe** wählen.
2. **Alt-Taste** drücken und gedrückt halten.
3. 4-stelligen Unicode-Codepoint erfassen.

- oder –

Wenn Sie ein kombinierendes Zeichen mit einem Basiszeichen kombinieren wollen: Basiszeichen erfassen und kombinierendes Zeichen erfassen. Das Basiszeichen und das kombinierende Zeichen werden zusammengeführt.

- Sonderzeichen erfassen mit dem Pop-Over-Menü

Wenn Sie mit der normalen Texteingabe Sonderzeichen erfassen wollen, können Sie das Pop-Over-Menü nutzen. Um das Pop-Over-Menü nutzen zu können, muss die Funktion Tastenwiederholung deaktiviert sein.

1. Taste des Zeichens gedrückt halten, das Teil des diakritischen Zeichens ist. Das Pop-Over-Menü wird angezeigt mit möglichen Sonderzeichen in der aktuellen Sprachauswahl.

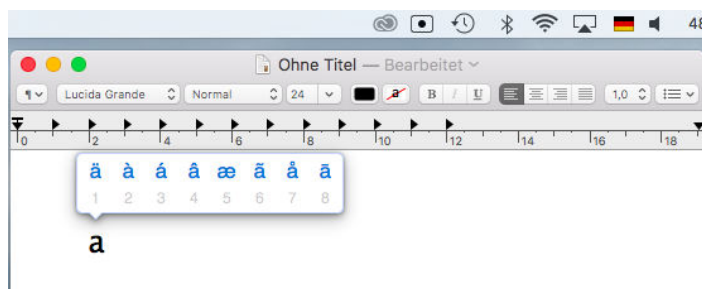


Bild 4: Sonderzeichen erfassen mit dem Pop-Over-Menü unter MacOS

2. Auf das Sonderzeichen klicken, das Sie erfassen wollen.

- oder –

Auf die Taste mit der Nummer drücken, die unter dem gewünschten Sonderzeichen angezeigt wird.

- E.7.8.** Welche Zeichen stammen aus welchen Länderalphabeten? Mit dieser Information können Funktionen implementiert werden, die Benutzern nur die erforderlichen Zeichen anzeigen und so die Anzeige nicht überfrachten und falsche Eingaben ähnlicher Zeichen reduzieren.

In den digitalen Anhängen latinchars.xml und 2018-11-15_din-spec-91379.xlsx, siehe Abschnitt C, befinden sich entsprechende Hinweise, die von Algorithmen verwendet werden können.

In der Excel-Datei wird in der Spalte „ns1:transliteration“ vollständig aufgelistet, ob ein Zeichen das Ergebnis einer bestimmten ISO-Transliteration sein kann. In der Spalte „ns1:countries“ werden die ISO Country Codes der Länder aufgelistet, in denen ein Zeichen üblicherweise verwendet wird. Für Zeichen aus dem weit verbreiteten Legacy Zeichensatz ISO 8859-15 ist die Aufzählung der ISO Country Codes nicht vollständig. Diese Zeichen sollten ggf. für alle Länder zur Auswahl gestellt werden. In der Spalte „ns1:iso-8859“ ist erkennbar, ob ein Zeichen aus dem Legacy-Zeichensatz stammt.

In der XML-Datei sind dieselben Informationen enthalten wie in der Excel-Datei, aber für Algorithmen leichter zugänglich.

Anhang F. Legacy Zeichensätze

Die Legacy-Zeichensätze ISO-8859-1, ISO-8859-15 und Windows-1252 waren Grundlagen der Betrachtungen bei der Zusammenstellung der Zeichen der DIN SPEC. Allerdings wurden die nachfolgend genannten Zeichen aufgrund einer fachlichen Bewertung nicht aufgenommen:

- a. Es werden grundsätzlich keine Steuerzeichen aufgenommen, also Zeichen der *General Kategorie Cc*. Die einzigen Ausnahmen sind die in Tabelle 8 genannten Zeichen *Character Tabulation U+0009*, *Line Feed U+000A* und *Carriage Return U+000D*.
- b. Es werden keine typografischen Zeichen aufgenommen. Dies betrifft folgende drei Zeichen des Legacy-Zeichensatzes Windows-1252:

CP 1252	Unicode	Zeichen	Hinweis
0x95	U+2022	Bullet	
0x96	U+2013	En Dash	Verwende U+002D (<i>Hyphen-Minus</i> : -)
0x97	U+2014	Em Dash	

- c. Der bedingte Trennstrich (*Soft Hyphen*, 0xAD entsprechend U+00AD) wurde nicht aufgenommen.

Anhang G. String.Latin in der Verwaltung

LA8 Passport der Bundesdruckerei

Die erste Festlegung einer auf die Bedarfslage der öffentlichen Verwaltung Deutschlands zugeschnittenen Teilmenge des Unicode-Standards war der von der Bundesdruckerei herausgegebene Zeichensatz „LA8-Passport“. Er diente der Erstellung hoheitlicher Dokumente auf Basis der Bestellungen der Pass- und Ausweisbehörden. Der Zeichensatz wurde nach Anforderung neuer lateinischer Zeichen der Pass- und Ausweisbehörden permanent angepasst.

Mit insgesamt 409 Zeichen in der letzten Aktualisierung ging er deutlich über das Leistungsvermögen der meisten Verfahren zur Führung von Pass-, Ausweis- und Melderegistern hinaus. Insbesondere wurde die Grenze von 256 Zeichen des in Europa weit verbreiteten ISO 8859-15 Zeichensatzes überschritten. Es bedurfte besonderer Software (Zeichentabellen in den Fachverfahren) zur Erfassung der Zeichen im Rahmen der Bestellung hoheitlicher Dokumente.

LA8-Passport war kein verfahrensübergreifender Standard. Er war speziell für den Zweck der Datenübermittlung für die Erstellung hoheitlicher Dokumente entwickelt worden. Die Tatsache, dass der auf dem Pass bzw. Ausweis gedruckte Name nicht immer mit den in elektronisch geführten dezentralen Registern gespeicherten Namen übereinstimmte, wurde als nach dem damaligen Stand der Technik unvermeidbar akzeptiert. Der Zeichensatz LA8-Passport wurde am 1. November 2012 durch String.Latin 1.1 in der Bundesdruckerei ersetzt.

Von den 409 Zeichen des LA8-Passport sind alle mit Ausnahme von 12 Ligaturen im normativen Bereich der DIN SPEC 91379 enthalten. Die zwölf Ausnahmen betreffen Zeichen wie beispielsweise NJ: LATIN CAPITAL LETTER NJ (U+01CA). Ligaturen sind in der DIN SPEC 91379 nicht enthalten, da sie nur typografischen Zwecken dienen, und die Komplexität des Vergleichs von Zeichenketten weiter erhöhen würden.

Elektronische Register im Personenstandswesen

Das ursprüngliche Motiv für die Festlegung einer verfahrensübergreifenden, auf die Bedarfslage der öffentlichen Verwaltung Deutschlands zugeschnittenen Teilmenge des Unicode Standards war die Entscheidung des Personenstandswesens zur Umstellung auf elektronische Registerführung im Jahr 2007. Da ein elektronischer Datenaustausch zwischen Personenstands- und Meldebehörden sowie indirekt auch Ausländerbehörden vorgesehen war, erfolgte eine Analyse der Vorgaben für die Registerführung dieser Bereiche. Dabei wurden Unterschiede bereits im Zeichensatz festgestellt.

Der für die Registerführung im Meldewesen einschlägige „Datensatz für das Meldewesen (DSMeld)“ verpflichtete IT Verfahren zur Unterstützung von insgesamt 59 Buchstaben: A...Z, a...z, ÄÖÜ, äöü und ß. Damit war ein Mindeststandard festgelegt. Die Unterstützung weiterer Zeichen war zulässig, bedurfte aber bilateraler Vereinbarungen. Die Existenz des von der Bundesdruckerei erstellten Zeichensatzes LA8 Passport beförderte solche Vereinbarungen, führte aber im Ergebnis nicht dazu, dass dieser umfangreiche Zeichensatz in allen bundesdeutschen Melderegistern umgesetzt worden wäre. Die fortschrittlichen IT Verfahren des Meldewesens hatten ISO 8859-15 umgesetzt und lagen damit oberhalb des Mindeststandards, aber unterhalb von LA8 Passport. Dabei ist zu berücksichtigen, dass erst ab dem 1. November 2007 (Einführung der Fingerabdrücke im Chip des Passes) alle Pass- und Ausweisbehörden die Antragsdaten digital übermittelten. Vorher war es im Papierverfahren durchaus üblich, dass Zeichen per Hand im Antrag nachgezeichnet wurden.

Im Ausländerzentralregister (AZR) wurden ebenfalls nur die wenigen, im Meldewesen verbindlich vorgegebenen Zeichen unterstützt. Über andere große IT Verfahren und Register der öffentlichen Verwaltung (Rentenversicherung, Kraftfahrt-Bundesamt, Finanzverwaltung, Bundeszentralregister etc.) machte die damalige Untersuchung keine präzisen Aussagen, es wurden jedoch analoge Verhältnisse wie im Meldewesen angenommen.

Für die neu einzurichtenden Personenstandsregister war jedoch die Unterstützung „aller lateinischen Zeichen“ vorzusehen. Namen mit diesen Zeichen sollten im Rahmen personenstandsrechtlicher Sachverhalte (Mitteilung der Geburt, der Heirat etc.) elektronisch an das Meldewesen, und in der Folge an andere Register übermittelt werden.

Beginn der Entwicklung (2008)

Aus den rechtlichen Vorgaben zur Schreibweise der Namen im Personenstandswesen folgte die Notwendigkeit zur Präzisierung der Formulierung „alle lateinischen Zeichen“ mit dem Ziel einer technischen Umsetzung sowohl für die Registerführung, als auch die Übermittlung an andere Behörden. Auf Vorschlag der KoSIT (damals noch als OSCI-Leitstelle) hat der zuständige Arbeitskreis I der Innenministerkonferenz im Jahr 2008 beschlossen:

Beschluss des AK I zu TOP 2.3 der 115. Sitzung im Herbst 2008 (Auszug: Ziffer 2)

Der AK I nimmt den Beschluss der DSMeld-Arbeitsgruppe zur Einführung des Standards ISO/IEC 10646:2003 in UTF 8-Kodierung zum 1. Mai 2011 zur Kenntnis. Er bittet die Leitstelle, die Entwicklung einer Aufstellung des zulässigen Zeichenumfangs und einer einheitlichen Lösung für die Altdaten in das Projekt „Deutschland Online Standardisierung“ einzubringen.

Die „Aufstellung des zulässigen Zeichenumfangs“ im Sinne der Festlegung einer Teilmenge des Unicode Standards erfolgte anschließend zunächst als Anlage des Standards XMeld für Datenübermittlungen von Personenstands- an Meldebehörden. Zur Ermittlung dieser Teilmenge wurden herangezogen:

- DIN 66303:1986-11 – Allgemeine Referenz-Version des 8-Bit-Code (ARV8) sowie ISO 8859-15
- Der Vorschlag für einen „paneuropäischen lateinischen Zeichensatz“ in [KAP 2007].
- Der Zeichensatz LA8 Passport der Bundesdruckerei.

Allen Beteiligten war klar, dass der Sachverhalt nicht nur auf technischer, sondern auch auf rechtlicher Ebene geregelt werden sollte, und dass nicht nur die Innenverwaltung betroffen ist (das im Beschluss des AK I der IMK adressierte „Projekt Deutschland Online Standardisierung“ kann als Vorläufer des IT-Planungsrats angesehen werden).

Festlegung der Version 1.0 in der Innenverwaltung (2011)

Am 30. Juni 2010 hat das BMI im Bundesanzeiger (eBAnz AT74 2010 B1) die „Bekanntmachung über die Herausgabe der Liste der im Meldewesen für die Registerführung und Datenübermittlungen und im Personenstandswesen für die elektronische Beurkundung und Datenübermittlung zulässigen lateinischen Zeichen (Buchstaben, Ziffern) und Sonderzeichen“ veröffentlicht. Es wurde festgelegt, dass „der Datentyp xoev:String.Latin sowohl für die Registerführung und für Datenübermittlungen im Meldewesen als auch für die elektronische Beurkundung und Datenübermittlung im Personenstandswesen ab 1. November 2011 zu verwenden [ist].“

Festlegung der Version 1.1 in der Innenverwaltung (2012)

Am 8. Februar 2012 hat das BMI im Bundesanzeiger (eBAnz AT24 2012 B1) die „Bekanntmachung über die Anwendung des Standards Lateinische Zeichen in Unicode und die Identifikation von Personen in automatisiert geführten Registern für die Bereiche Meldewesen, Personenstandswesen und Ausländerwesen“ veröffentlicht. Es wurde bestimmt:

- Für die elektronische Registerführung und Datenübermittlung im Meldewesen und Personenstandswesen sowie für die Datenübermittlung im Ausländerwesen ist ab 1. November 2012 der Standard „Lateinische Zeichen in Unicode“ in der Version 1.1.1 zu verwenden.
- Um eine sichere Identifikation von Personen zu ermöglichen, sind in den Bereichen Meldewesen, Personenstandswesen und Ausländerwesen ab dem 1. November 2012 die Vorgaben für Identifikationsverfahren aus dem Bericht der Projektgruppe Standard des Arbeitskreis I der Innenministerkonferenz „Umstellung auf lateinische Zeichen in Unicode – Vorgaben für Identifikationsverfahren“ anzuwenden.

Ressortübergreifende Festlegung der Version 1.1 (2014)

Der IT-Planungsrat hat im Jahr 2012, im Rahmen seiner 7. Sitzung mit der Entscheidung 2012/05 [https://www.it-planungsrat.de/SharedDocs/Sitzungen/DE/2012/Sitzung_7.html?pos=5], erstmals festgelegt, dass er die Festlegung eines Zeichensatzes durch einen fachunabhängigen IT-Interoperabilitätsstandard im Sinne des §3 des IT-Staatsvertrags für erforderlich hält.

Entscheidung 2012/05

1. Der IT-Planungsrat begrüßt die Herausgabe des Standards „Lateinische Zeichen in UNICODE“ (Anlage 1) durch die Koordinierungsstelle für IT-Standards.
2. Der IT-Planungsrat stellt fest, dass es sich um einen fachunabhängigen IT-Interoperabilitätsstandard im Sinne des § 3 des IT-Staatsvertrages handelt. Er empfiehlt dessen Umsetzung in allen IT-Verfahren der öffentlichen Verwaltung.
3. Der IT-Planungsrat bittet die KoSIT und das BMI, die Bestrebungen zur Standardisierung auf europäischer Ebene auch weiterhin zu unterstützen.
4. Der IT-Planungsrat stellt fest, dass der Bedarf zur Festlegung eines Zeichensatzes, der von den IT-Verfahren der öffentlichen Verwaltung bei Registerführung und Datenübermittlung unterstützt werden muss, zusammen mit dem Standard „Lateinische Zeichen in UNICODE“ auf die Standardisierungsagenda aufzunehmen ist.

Die weitere Bearbeitung erfolgte dann im Rahmen der für den IT-Planungsrat geführten Standardisierungsagenda. Im Jahr 2014 waren die Arbeiten abgeschlossen so dass der IT-Planungsrat im Rahmen seiner 13. Sitzung mit seiner Entscheidung 2014/04 [https://www.it-planungsrat.de/SharedDocs/Sitzungen/DE/2014/Sitzung_13.html?pos=4] die Umsetzung binnen drei Jahren beschlossen hat. Der Beschluss wurde am 12.3.2014 gefasst, so dass die Verpflichtung zur Umsetzung von String.Latin 1.1 seit dem 12.3.2017 für alle IT-Verfahren gilt, die dem bund-länderübergreifenden Datenaustausch oder dem Datenaustausch mit Bürgern und Wirtschaft dienen.

Beschluss 2014/04

1. Unter Bezug auf § 1 Abs. 1 Satz 1 Nr. 2 des Vertrages über die Errichtung des IT-Planungsrats und über die Grundlagen der Zusammenarbeit beim Einsatz der Informationstechnologie in den Verwaltungen von Bund und Ländern (IT-Staatsvertrag) beschließt der IT-Planungsrat die verbindliche Anwendung des Interoperabilitätsstandards „Lateinische Zeichen in UNICODE“ als Mindeststandard.
2. Für IT-Verfahren, die dem bund-länderübergreifenden Datenaustausch oder dem Datenaustausch mit Bürgern und Wirtschaft dienen, werden folgende Fristen für die Konformität laut Anlage 1 festgelegt:
 - mit Beschlussfassung - für IT-Verfahren, die neu aufgebaut oder in wesentlichem Umfang überarbeitet werden,
 - drei Jahre nach Beschlussfassung - für andere IT-Verfahren.
3. Die Mitglieder des IT-Planungsrats tragen in ihrer jeweiligen Gebietskörperschaft dafür Sorge, dass, sobald möglich, sämtliche IT-Verfahren konform zu diesem Standard sind, wenn nicht zwingende fachliche oder wirtschaftliche Gründe dagegen sprechen.
4. Der Standard „Lateinische Zeichen in UNICODE“ wird im Auftrag des IT-Planungsrats von der Koordinierungsstelle für IT-Standards (KoSIT) herausgegeben. Der Standard ist im Bundesarchiv, Potsdamer Straße 1, 56075 Koblenz, für jedermann zugänglich und archivmäßig gesichert niedergelegt.

5. Der Standard und darauffolgende Änderungen werden im Bundesanzeiger bekannt gemacht.

Die in den Ziffern 2 und 3 angesprochene „Konformität von IT-Verfahren“ ist wie folgt festgelegt: „Ein IT-Fachverfahren ist konform zum Standard Lateinische Zeichen in UNICODE, wenn jedes der im Standard Lateinische Zeichen in UNICODE durch seinen Codepoint bzw. seine Codepoints bezeichnete Zeichen in dem IT-Fachverfahren erfasst, verarbeitet, gespeichert, übermittelt und gedruckt werden kann.“ Diese Festlegung erwies sich im Nachhinein als zu unbestimmt. Die Betreiber großer IT-Verfahren bemängelten, dass nicht erkennbar sei, welche Datenfelder zwingend umgestellt werden müssen um Konformität herzustellen. Die mögliche Bandbreite reicht von „Alle Datenfelder für Namen natürlicher Personen“ bis hin zu „sämtliche Datenfelder des jeweiligen IT-Verfahrens“.

Diese Schwierigkeiten führten bei der anschließenden Erstellung der DIN SPEC 91379 zur Formulierung von „Namen im weiteren Sinne“, die im ersten Absatz des Abschnitts 1 - *Anwendungsbereich* wie folgt bestimmt werden: „Namen im weiteren Sinne sind Namen natürlicher Personen gemäß der Vorgaben des Personenstandsrechts ebenso wie Namen juristischer Personen, Namen von Produkten, Orten und Straßen sowie Titel von Dokumenten oder Gesetzen. Namen im weiteren Sinne können konkrete Objekte bezeichnen, aber auch virtuelle Konstrukte wie Produktgruppen oder Musikstile.“ In Verbindung mit den Festlegungen zur Konformität in Abschnitt 4.1 sollten die genannten Schwierigkeiten nunmehr beseitigt sein.

Überführung in die nationale Normung (2017)

Im Rahmen seiner 24. Sitzung hat der IT-Planungsrat mit der nachfolgenden Entscheidung 2017/41 die Überführung in die nationale und europäische Normung beschlossen.

Entscheidung 2017/41 des IT-Planungsrats

1. Der IT-Planungsrat begrüßt die Einrichtung einer Fachgruppe „String.Latin“ unter Federführung der KoSIT. Er fordert seine Mitglieder auf dafür zu werben, sich an der Fachgruppe zu beteiligen. Er bittet die Fachgruppe, ihm Vorschläge zu unterbreiten, die geeignet sind, die Umsetzung des Standards zu befördern.
2. Der IT-Planungsrat begrüßt das Ziel, den Standard „Lateinische Zeichen in UNICODE“ in die europäische Normung einzubringen. Er bittet die Koordinierungsstelle für IT-Standards (KoSIT) und die Mitglieder der in Ziffer 1 bezeichneten Fachgruppe, entsprechende Aktivitäten beim Deutschen Institut für Normung (DIN) aktiv zu unterstützen. Er unterstützt die Absicht, die Version 1.2 des Standards „Lateinische Zeichen in UNICODE“ in Form einer DIN SPEC zu erstellen.
3. Er wird hierfür einen Betrag von bis zu 50.000 EUR zweckgebunden in der Finanzplanung 2018 bereitstellen.
4. Der IT-Planungsrat begrüßt die Entscheidung des EESSI Projektes zur Verwendung des Standards „Lateinische Zeichen in UNICODE (String.Latin)“. Er bittet die an dem Projekt beteiligten deutschen Behörden und Organisationseinheiten um aktive Unterstützung durch Beteiligung an der unter Ziffer 1) bezeichneten Fachgruppe und durch Mitarbeit bei der in Ziffer 2) dargelegten Einbringung in die europäische Normung.

Auf der Basis dieser Beschlusslage hat die KoSIT die Erstellung einer DIN SPEC beim Deutschen Institut für Normung e.V. initiiert. Die inhaltlichen Arbeiten fanden im Jahr 2018 statt. Eine Entwurfsfassung zur Kommentierung durch die Öffentlichkeit wurde im Herbst 2018 veröffentlicht. Basierend auf den Kommentaren zur Entwurfsfassung wurden die inhaltlichen Arbeiten Ende 2018 abgeschlossen. Am 8. März 2019 wurde die DIN SPEC 91379 vom Beuth Verlag veröffentlicht.

Festlegung der DIN SPEC 91379 (2019)

Im Rahmen seiner 28. Sitzung am 12. März 2019 hat der IT-Planungsrat den Beschluss 2019 / 16 [https://www.it-planungsrat.de/SharedDocs/Sitzungen/DE/2019/Sitzung_28.html?pos=16] gefasst:

Beschluss 2019 / 16 des IT-Planungsrats

1. Der IT-Planungsrat begrüßt die Herausgabe der DIN SPEC 91379 durch das Deutsche Institut für Normung e.V. als Weiterentwicklung des in seinem Auftrag von der KoSIT herausgegebenen Standards „Lateinische Zeichen in Unicode“.
2. Er gibt die DIN SPEC 91379 für die Verarbeitung von Namen (im weiteren Sinne, also Namen natürlicher und juristischer Personen sowie Namen von Objekten wie beispielsweise Orten, Straßen oder Produkten) vor.
3. Er bittet seine Mitglieder, die Umsetzung der DIN SPEC 91379 für IT-Verfahren, die dem Bund-Länder übergreifenden Datenaustausch oder dem Datenaustausch mit Bürgern und Wirtschaft dienen, unverzüglich zu beginnen und der Fachgruppe „String.Latin“ jeweils zum 1. Februar 2020 und zum 1. Februar 2022 zum Stand der Umsetzung zu berichten. Für diese IT-Verfahren ist jeweils zu dokumentieren:
 - a. der erreichte Stand
 - b. das weitere Vorgehen zum Erreichen der Konformität
 - c. der Zieltermin der vollständigen Konformität zur DIN SPEC 91379
 - d. Risiken, ggf. Probleme und Gründe für Verzögerungen
4. Der IT-Planungsrat wird in seiner 30. Sitzung [am 24. Oktober 2019] den Zeitpunkt festlegen, bis zu dem alle IT-Verfahren, die dem Bund-Länder übergreifenden Datenaustausch oder dem Datenaustausch mit Bürgern und Wirtschaft dienen, konform zur DIN SPEC 91379 sein müssen.
5. Er begrüßt die geplante Weiterentwicklung der DIN SPEC zu einer DIN-Norm und anschließend einer CEN-Norm. Er bittet die KoSIT und die Mitglieder der Fachgruppe „String.Latin“, sich aktiv an dieser Fortschreibung des Standards zu beteiligen.

Der obige Beschluss wurde durch die Entscheidung 2019/53 [https://www.it-planungsrat.de/SharedDocs/Sitzungen/DE/2019/Sitzung_30.html?pos=15] der 30. Sitzung ergänzt:

In Ergänzung seiner Entscheidung 2019/16 legt der IT-Planungsrat fest, dass alle IT-Verfahren, die dem Bund-Länder übergreifenden Datenaustausch oder dem Datenaustausch mit Bürgern und Wirtschaft dienen, spätestens zum 1. November 2024 konform zur DIN SPEC 91379 sein müssen.

Durch diese beiden Beschlüsse wurde klargestellt:

- Die DIN SPEC 19379 ist der Nachfolger von String.Latin 1.1. (Aus diesem Grund wird sie manchmal informell als „String.Latin 1.2“ bezeichnet.)
- IT-Verfahren der öffentlichen Verwaltung, die dem Bund-Länder übergreifenden Datenaustausch oder dem Datenaustausch mit Bürgern und Wirtschaft dienen, müssen die DIN SPEC 91379 **bis spätestens zum 1. November 2024** für die Verarbeitung von Namen (im weiteren Sinne) umsetzen.

Damit wird die Festlegung der Konformität von IT Verfahren zur DIN SPEC in Abschnitt 4.1 aufgegriffen. Dies ist ein wesentlicher Fortschritt gegenüber der Beschlusslage zu String.Latin 1.1, denn dort war unklar geblieben, welche konkreten Maßnahmen erforderlich sind um die Konformität von IT-Verfahren herzustellen.

- Der Weg der Normung soll konsequent weiterverfolgt werden, indem die DIN SPEC 91379 zunächst in eine nationale DIN Norm und anschließend (ggfs. im Wege eine *fast-track* Verfahrens) in eine CEN-Norm überführt werden soll.

Literaturhinweise und Weblinks

Die DIN SPEC enthält in ihrem normativen Teil alle Zeichen, die Transliterationsergebnis einer der hier aufgeführten ISO-Normen sein können. Für diese stellt die DIN SPEC die Kompatibilität sicher. Wenn eine der ISO-Normen weiterentwickelt wird, entsteht Prüfbedarf für die DIN SPEC.

[AKI Identifikation] *Umstellung auf Lateinische Zeichen in Unicode – Vorgaben für Identifikationsverfahren*. . 17. Januar 2012. Online erhältlich bei der KoSIT [https://www.xoev.de/die_standards/lateinische_zeichen_in_unicode/umstellung_auf_lateinische_zeichen_in_unicode__vorgaben_fuer_identifikationsverfahren-4811].

[Confusables] *Unicode Utilities: Confusables*. Demonstration von Sicherheitsrisiken aufgrund leicht verwechselbarer Zeichen auf einer Webseite von Unicode [<https://unicode.org/cldr/utility/confusables.jsp>].

[Dejavu Fonts] *Dejavu Fonts bei Github*. Dejavu Fonts, erhältlich bei GitHub [<https://github.com/dejavu-fonts>].

[FAQ Normalization] *Frequently Asked Questions - Normalization*. Ausschließlich Online auf der Webseite von Unicode [<https://unicode.org/faq/normalization.html>].

[ICAO 9303-3] *Machine Readable Travel Documents – Part 3: Specifications common to all MRTDs*. . 7th Edition 2015. Im Internet erhältlich bei ICAO [https://www.icao.int/publications/Documents/9303_p3_cons_en.pdf].

[ISO 11940] *Information and documentation — Transliteration of Thai*. Siehe Eintrag bei Wikipedia [https://en.wikipedia.org/wiki/ISO_11940].

[ISO 15915] *Information and documentation — Transliteration of Devanagari and related Indic scripts into Latin characters*. Siehe Eintrag bei Wikipedia [https://en.wikipedia.org/wiki/ISO_15919].

[ISO 233] *Information and documentation — Transliteration of Arabic characters into Latin characters — Part 3: Persian language — Simplified transliteration*. Siehe Eintrag bei Wikipedia [https://en.wikipedia.org/wiki/ISO_233].

[ISO 233-3] *Information and documentation — Transliteration of Arabic characters into Latin characters — Part 3: Persian language — Simplified transliteration*. Siehe Eintrag bei Wikipedia [https://en.wikipedia.org/wiki/ISO_233].

[ISO 259] *Documentation — Transliteration of Hebrew characters into Latin characters*. Siehe Eintrag bei Wikipedia [https://en.wikipedia.org/wiki/ISO_259].

[ISO 259-2] *Information and documentation — Transliteration of Hebrew characters into Latin characters — Part 2: Simplified transliteration*. Siehe Eintrag bei Wikipedia [https://en.wikipedia.org/wiki/ISO_259].

[ISO 3602] *Documentation — Romanization of Japanese (kana script)*. Siehe Eintrag bei Wikipedia [https://en.wikipedia.org/wiki/ISO_3602].

[ISO 7098] *Information and documentation — Romanization of Chinese*. Siehe Eintrag bei Wikipedia [https://en.wikipedia.org/wiki/ISO_7098].

[ISO 843] *Information and documentation — Conversion of Greek characters into Latin characters*. Siehe Eintrag bei Wikipedia [https://en.wikipedia.org/wiki/ISO_843].

[ISO 8859-15] *Information technology — 8-bit single-byte coded graphic character sets — Part 15: Latin alphabet No. 9*. Siehe Eintrag bei Wikipedia [https://en.wikipedia.org/wiki/ISO/IEC_8859-15].

-
- [ISO 9] *Information and documentation – Transliteration of Cyrillic characters into Latin characters — Slavic and non-Slavic languages*. Siehe Eintrag bei Wikipedia [https://en.wikipedia.org/wiki/ISO_9].
- [ISO 9984] *Information and documentation — Transliteration of Georgian characters into Latin characters*. Siehe Eintrag bei Wikipedia [https://en.wikipedia.org/wiki/ISO_9984].
- [ISO 9985] *Information and documentation — Transliteration of Armenian characters into Latin characters*. Siehe Eintrag bei Wikipedia [https://en.wikipedia.org/wiki/ISO_9985].
- [IT-Staatsvertrag] *Vertrag über die Errichtung des IT-Planungsrats und über die Grundlagen der Zusammenarbeit beim Einsatz der Informationstechnologie in den Verwaltungen von Bund und Ländern*. Vertrag zur Ausführung von Artikel 91c GG. Online erhältlich bei buzer.de [<https://www.buzer.de/IT-Staatsvertrag.htm>].
- [KAP 2007] *Zeichen setzen für Europa*. Der Gebrauch europäischer lateinischer Sonderzeichen in der deutschen Öffentlichkeit. Bernd Kappenberg. Online erhältlich bei Networx Nr. 49 [<https://www.mediensprache.net/de/networx/docs/networx-49.aspx>]. Rev 2007-02-18.
- [Liberation Fonts] *Liberation Fonts bei GitHub*. Liberation Fonts, erhältlich bei Github [<https://github.com/liberationfonts>].
- [Noto Fonts] *Google Noto Fonts*. Beautiful and free fonts for all languages. Online erhältlich bei Google [<https://www.google.com/get/noto/>].
- [Open Source Fonts] *Open-source Unicode typefaces*. Ausschließlich online erhältlich bei Wikipedia [https://en.wikipedia.org/wiki/Open-source_Unicode_typefaces].
- [String Latin 1.1] https://www.icao.int/publications/Documents/9303_p3_cons_en.pdf. . Version 1.1.1 vom 27.1.2012. Von der KoSIT herausgegeben im Auftrag des IT-Planungsrats. Im Internet erhältlich bei der KoSIT [http://xoev.de/latinchars/1_1/latinchars.pdf].
- [Unicode 12] *The Unicode Standard Version 12.0 — Core Specification*. . Im Internet erhältlich von Unicode [<https://www.unicode.org/versions/Unicode12.0.0/UnicodeStandard-12.0.pdf>].
- [Unicode 15] *Unicode Standard Annex #15 - Unicode Normalization Forms*. Im Internet erhältlich von Unicode [<http://unicode.org/reports/tr15/>].
- [Unicode 34] *Unicode Standard Annex #34 - Unicode named character sequences*. Im Internet erhältlich von Unicode [<http://www.unicode.org/reports/tr34/tr34-23.html>].
- [Unicode 42] *Unicode Standard Annex #42 - Unicode character database in XML*. Im Internet erhältlich von Unicode [<http://unicode.org/reports/tr42/>].
- [Unicode 9] *The Unicode Standard Version 9.0 — Core Specification*. . Im Internet erhältlich von Unicode [<https://www.unicode.org/versions/Unicode9.0.0>].
- [Unicode TR 36] *Unicode Technical Report #36 - Unicode Security Considerations*. Im Internet erhältlich von Unicode [<http://unicode.org/reports/tr36/>].
- [W3C CHARMOD] *Character Model for the World Wide Web 1.0: Fundamentals*. . 15 February 2005. W3C Recommendation [<https://www.w3.org/TR/charmod>].
- [W3C CHARMOD-NORM] *Character Model for the World Wide Web 1.0: String Matching*. . 4 February 2019. W3C Working Group Note [<https://www.w3.org/TR/charmod-norm/>].